Tampere University

Tomi Palovuori

# THE ANALYSIS AND USE OF MOTOR VEHICLE TELEMETRY DATA

Master of Science Thesis
Faculty of Computing and Electrical Engineering
Examiners: David Hästbacka
Petri Kannisto
June 2022

# ABSTRACT

Tomi Palovuori: The analysis and use of motor vehicle telemetry data
Master of Science Thesis
Tampere University
Master's Degree in Computer Science
June 2022

---

In order to operate a vehicle, information about the vehicle must be communicated to the operator, as it is difficult or impossible to determine otherwise. Information such as vehicle speed, remaining fuel or whether seatbelts are on can be measured and transmitted electrically. Some measurements can be specific to a vehicle type or require processing with other metrics to provide value. The measurements can be displayed immediately or stored for a longer period of time before they are sent for a longer time period analysis.

Besides the immediate need to access vital operating metrics, there are also other uses for gathered data. Analyzing measurements from a longer period of time can show trends in time. Combining various metrics can also provide a tangible metric for more abstract concepts, such as vehicle efficiency, convenience or safety. Data is also more valuable if gathered from a larger number of similar vehicles. A larger business can use a larger dataset to obtain good comparison points and a robust average for future estimates.

The value derived from vehicle telemetry in businesses appears in different forms. Most notably, business analytics can be applied to reduce vehicle downtime and predict future profits and costs. The performance of hired drivers can also be compared to each other to improve performance by rewarding good driving practices. Measurements a vehicle makes can also be used to discover indicators for parts wear and predict future failures. Vehicle telemetry is not only limited to private use, as services such as vehicle live location are brought to consumers as well through different applications.

In the future, more data is required for more complicated applications. For autonomous vehicles, the vehicle will require measurements an operator would normally make themselves, such as a visual of other vehicles. Besides the operating capacity, the ability to detect wear and smaller malfunctions within the vehicle itself is imperative, as there might be no person to notice and report them. A concrete stepping stone for this is the ability to track passengers within the vehicle. For public transportation, the vehicle must know when passengers are done entering and leaving. If the amount of people is measured, it is also possible to deduce the number of people inside the vehicle at a given point in time. This information can be used by a business to measure the number of passenger miles a vehicle provides, or by consumers in order to avoid crowded transportation vehicles.

Keywords: Telemetry, Vehicle telemetry, Transportation, Data science, Data analysis, Data handling

# TIIVISTELMÄ

Tomi Palovuori: Ajoneuvotelemetriadatan käyttö ja analyysi
Diplomityö
Tampereen yliopisto
Tietotekniikan DI-tutkinto
Kesäkuu 2022

Ajoneuvon turvallista käyttämistä varten tarvitaan informaatiota sen sisäisestä toiminnasta. Ajoneuvon nopeus, polttoaineen määrä ja turvavöiden kytkentä voidaan mitata shköisesti ja viestiä kuljettajalle. Jotkin mittaukset ovat rajoitettu tiettyihin ajoneuvotyyppeihin, ja mittauksilla voidaan tehdä monimutkaisempia laskuja arvokkaan tiedon saamiseksi. Mittaukset voidaan näyttää ja unohtaa välittömästi, tai niitä voidaan säilöä pidempi aika, jolloin mahdollistetaan pidemmän ajanjakson mittausten analyysi.

Turvallisen ajamisen mahdollistamisen lisäksi mittauksia voidaan tehdä myös muista osista muihin tarkoituksiin. Pidemmän ajanjakson analysointi voi osoittaa nousevia tai laskevia trendejä mitattavilla osa-alueilla. Yhdistelemällä useita mittaustuloksia voidaan saada myös konkreettisia mittaraita abstrakteille osa-alueille, kuten ajamisen tehokkuus tai turvallisuus. Mittausdatan arvokkuus myös kasvaa yhdistäessä tuloksia useista samankaltaisita ajoneuvoista. Isompi yritys voi hyödyntää isoa määrää dataa saadakseen tarkkoja vertailutuloksia ja arvioita tulevaisuuteen.

Ajoneuvotelemetriasta saatava rahallinen hyöty yrityksille syntyy useiden erilaisten sovellusten kautta. Isoinpana sovelluksena on liiketoiminnan analysointi, jonka avulla voidaan parantaa ajoneuvojen käytön tehokkuutta ja ennustaa tulevia tuloja ja menoja. Yrityksen palkkaamien kuskien ajotehokkuutta voidaan myös seurata datasta. Hyviä ajotapoja voidaan palkita rahallisesti, jotta kuskien ajotehokkuus nousisi. Ajoneuvotelemetrian mittaustuloksista voidaan etsiä myös viitteitä ajoneuvon osien kulumiseen tai hajoamiseen, jotta osien merkittävä hajoamisriski voidaan huomata ajoissa. Myös kuluttajat voivat hyötyä ajoneuvotelemeriasta sovelluksien, kuten ajoneuvojen sijainnin seurannan kautta.

Tulevaisuudessa ajoneuvojen keräämän telemetriadatan määrä tulee kasvamaan, jotta monimutkaisempia sovelluksia voidaan saada käyttöön. Itseohjautuvat autot tarvitsevat tietoonsa kaiken sen, mitä kuski tarvitsee ajaakseen, esimerkiksi näköyhteys muihin ajoneuvoihin. Ajotuntuman puuttuessa tarvitaan myös tarkempaa tietoa osien toiminnasta, sillä kuski ei ole huomaamassa ajamisessa esiintyviä pieniä ongelmia. Esimerkki tarvittavasta mittauksesta on matkustajien liikkuminen ajoneuvon sisällä. Julkisessa liikenteessä bussin, junan tai raitiovaunun tulee tietää, milloin on turvallista sulkea ovet ja jatkaa matkaa. Jos ihmisten liikkumista ajoneuvoon ja sieltä ulos voidaan mitata, sisällä olevien matkustajien kokonaismäärä voidaan selvittää matkoilla. Tätä tietoa voidaan hyödyntää sekä ajoneuvon kuljetustehokkuuden laskemiseen yksityisesti, että täysien ajoneuvojen välttelemiseen yksilötasolla.

Avainsanat: Telemetria, Ajoneuvotelemetria, Liikenne, Datan analysointi, Tiedonkäsittely

# PREFACE

This thesis is the result of varying intensity work over the period of 18 months. Therefore I would first like to thank the main examiner of this thesis, David Hästbacka, for bearing with it and me. Similarly, my superiors at Knowit Finland, Jutta Luhtalampi and Tapani Paajoki, deserve a thanks for their patience. As this is the final piece of my studies towards a master of science, a general thank you is in order for the entire University of Tampere and its personnel. Finally, I would like to thank the people of Tupsula, without whom this thesis and my studies would have been completed years earlier.

Tampere, 30th June 2022

Tomi Palovuori

# CONTENTS

1. Introduction . . . 1
2. Telemetry . . . 3
  2.1 Definition . . . 3
  2.2 Storage . . . 4
  2.3 Transmission . . . 6
  2.4 Data reception and processing . . . 8
  2.5 Data analysis . . . 9
  2.6 Data visualization . . . 12
3. Vehicle Telemetry . . . 14
  3.1 Railway telemetry . . . 14
    3.1.1 History . . . 14
    3.1.2 Modern railways . . . 15
  3.2 Automobile telemetry . . . 16
    3.2.1 History . . . 16
    3.2.2 Automobile telemetry signals . . . 17
  3.3 Passenger car telemetry . . . 19
  3.4 Delivery telemetry . . . 22
  3.5 Transportation telemetry . . . 24
    3.5.1 Problem setting . . . 24
    3.5.2 Data use in transportation . . . 25
    3.5.3 Transportation metrics . . . 27
    3.5.4 Use of telemetry . . . 29
  3.6 Vehicle fleet telemetry . . . 34
4. Telemetry data applications . . . 37
  4.1 Private use of telemetry data . . . 37
    4.1.1 Business data analytics . . . 38
    4.1.2 Driver analytics . . . 40
    4.1.3 Predictive maintenance . . . 44
  4.2 Public use of telemetry data . . . 45
5. Future of vehicle telemetry . . . 48
  5.1 Autonomous vehicle telemetry . . . 48
  5.2 Public transportation fullness measuring . . . 50
6. Discussion . . . 52
  6.1 Measured signals . . . 52

6.2 Telemetry applications . . . . . . . . . . . . . . . . . . . . . . . . 53
6.3 Vehicle telemetry's future . . . . . . . . . . . . . . . . . . . . . . 54
7. Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| ABS | Anti-lock Braking System |
| AI | Artificial Intelligence |
| AWS | Amazon Web Services |
| CEO | Chief Executive Officer |
| EU | European Union |
| GPS | Global Positioning System |
| GPU | Graphical Processing Unit |
| IoT | Internet of Things |
| MPG | Miles Per Gallon |
| MTS | Martin Transportation Systems |
| RPM | Revolutions Per Minute |
| UK | United Kingdom |
| US | United States |
| WIM | Weigh-In-Motion |

# 1. INTRODUCTION

Ever since the last decade's artificial intelligence boom, it has become increasingly obvious that data is the limiting factor in machine learning. With graphical processing units, GPUs, advanced enough for complicated processing of enormous amounts of data, the shift has moved to the qualities of data itself. Improving the accuracy and expressiveness of data now takes precedence over increasing its quantity. Cleaning bad samples, outliers and irrelevant characteristics from the data brings back the understanding of data into data analysis. Identifying these is an entire problem on its own, though.

Understanding how gathering and analyzing data can offer value in a field is sometimes difficult. In the field of motorized vehicles, however, there are good examples. The most down-to-earth example is the odometer installed in almost every motorized vehicle. Odometer, also known as "mileage meter", gathers data in form of driven distance. This information is considered so valuable and expressive of the vehicle's condition that odometers are mandated by law, and tampering with them is a crime [1, p.11-15]. While checking the mileage on one's own car is simple, expanding the scope to an entire second-hand car shop, for example, is not as simple. Each car would have to be turned on, the meter then read and written down. The entire operation would also be repeated every time the mileages should be checked. The data from such inaccessible instruments, its recording and transmitting are called telemetry.

Telemetry is in many ways unique from traditional data. For it's recording, the amount of data collected between transmissions has to fit into the instrument. Adding additional measured metrics would require the installation of new sensors and is typically out of the question. Due to the inherent nature of telemetry, the data will arrive in chunks from each instrument. This introduces spikes in required processing power as well as differing amounts of lag. If recording malfunctions, it will take until the next transmission to spot it and all the data in between will be lost.

There are numerous use cases for telemetry data. Most importantly, it is required to operate any modern vehicle. The telemetry readings from various parts of a vehicle are collected and presented to the driver on the dashboard. For this data, it is imperative that delays are minimized and the recording has no breaks. The historical telemetry data from a long period of time has different applications. It can provide accurate details about vehicle operations, from location history to fuel consumption. This data can be used for business intelligence purposes, leading to more efficient operations with less unknown variance.

The goal of this thesis is to study vehicle telemetry and its applications. The actual measuring devices are left out of scope, and the measurements they make are assumed to be accurate enough for introduced purposes. Outside of this, the three key aspects this thesis covers are the variety of signals that vehicles measure, what and how modern applications of vehicle telemetry use these signals and what future applications and measurements could be made.

The three key aspects of vehicle telemetry are approached through a theoretical base as well as example cases. Through exploring the telemetry signals common in almost every vehicle and why they exist, a good starting point can be found. Continuing to more unique signals used in vehicles made for different purposes, key telemetry data can be found. How this data is implemented can then both be examined through the theoretical possibilities as well as companies that offer these implementations. As the exact manners of implementation are trade secrets, the theory will allow for approximation of the methods used. Future improvements in telemetry data use will happen in numerous fields, so the focus is narrowed to two inherently different example cases, one in which groundbreaking progress is made and the other already possible, merely lacking implementation.

This thesis is structured as follows: In Chapter 2, the theory behind telemetry data is explained. Telemetry specifically in the field of vehicles is explored in Chapter 3. Chapter 4 discusses modern applications of telemetry data in both public and private use with examples. In Chapter 5, the future of vehicle telemetry applications is viewed through two example cases. Finally, the key questions are discussed in Chapter 6 and the thesis is concluded in Chapter 7.

# 2. TELEMETRY

In this chapter, the common fundamentals for telemetry data are laid out. These fundamentals are paired with easy-to-understand examples for vehicle telemetry but are nevertheless common for many if not all telemetry applications. The splits into parts in various topics are made arbitrarily by the author in order to ease explanation and group areas to better reflect topics further in this thesis.

## 2.1 Definition

There are somewhat differing definitions for telemetry. According to Merriam-Webster, it is the science or process of telemetering data, which is either the measured distance of a distant object to an observer or an electric reading of a quantity transferred to a distant station [2] [3]. Cambridge dictionary states it as "the science or process of collecting information about objects that are far away and sending the information somewhere electronically" [4]. Both of these imply that the measured aspect has to be far away physically. However, since telemetry has become key in medical research as well as more intricate machine constructs, the current use of the term is a bit broader. Princeton Wordnet and the consensus of Wiktionary prefer "remote source" instead of "far away object" in their definitions [5] [6]. This wording includes cases where the measurements are made physically close, but in hard-to-access places. Or in this thesis' case, inconvenient places. When driving a car, it is much more convenient to use telemetry and read the gas meter than to open the hatch to the tank and measure it by oneself.

Telemetry by its definition can be split into three parts:

- Obtaining the data (measurement or its derivative)
- Storing the data (sometimes optional)
- Transmitting the data (electronically)

One of the earliest common cases of obtaining, storing and transmitting data was for trains. Train stations would collect data of trains (which station the train is at, time of departure), store it in records, and transmit it via telegraph to other stations. In time this has evolved to live tracking of all transportation. The benefits of this use of vehicle telemetry is apparent, from station timetables to airspace control.

Obtaining data for telemetry is rarely done by humans like in the train example. Generally, an electronic sensor will produce a reading of a measured quantity at set intervals. The data processing aspect begins once the measurement has been made and as such will not be focused on this thesis. The measurement might not be the data transferred, as edge computing might be made before transfer [7]. Anything from a simple AND-gate to complex behavioral pattern recognition can be applied to signals before they are stored and/or transmitted.

Storing the data is not always a necessary part of the pipeline from a sensor to transmission. Signals like the GPS location for live tracking should be transmitted as soon as possible. They can of course be kept by the receiver for analysis during long periods, but then it is no longer part of the telemetry pipeline. When the transmission connection is only formed periodically, the data collected between the transmission needs to be stored somewhere. Or kept stored safely "just in case", like in the black boxes in aircraft [8]. Data volumes, amounts and storage is further discussed in Chapter 2.2.

Electronically transmitting data is on the surface level the simplest part of the pipeline. Transmission can be made either wired or wireless. In wired transmission, there usually is no need for storage as the wired connection is constant. This is the case for most data relayed from the vehicle to its driver, such as speed, amount of gas and turn signals. Wired connections can also be made only periodically, such as when a vehicle is brought in for maintenance. Storage is in these cases essential. Wireless transmissions were originally made by radio for a long time, but can also be sent through the internet. As there are areas where transmissions cannot be made, either due to lack of range for radio or bad reception. The data must be stored or it is lost in unsuccessful transmissions. The transmission of data is discussed more in Chapter 2.3.

## 2.2 Storage

According to Moore's Law, the cost of memory halves every 18 months. While we are approaching a breaking point in that trend, the cost of memory for storing signals in smart intervals and precision is almost negligible. [9] A 32-bit single precision floating point measurement stored every second for a year, 31 536 000 times, takes almost exactly one gigabit of storage. Within a 50-dollar 1-terabyte hard drive, one could fit about 8 000 of these signals [10]. If the data is compressed to storage, then that hard drive would fit multiple times more measurements.

There is very little need for second-interval measurements that would be only uploaded once a year in vehicle telemetry. If measurement interval is short or precision is required, then usually lag is wanted to be kept to a minimum. As soon as measured history is transferred, it can usually be removed from storage. If the transmission is made upon request to multiple parties, it can be kept a for longer period, but even then history is eventually deemed unnecessary and removed. Therefore the volume of storage is rarely a limiting factor. What is more important is what needs to be stored. This is discussed in length in Chapter 3.

These are the main types of data to consider in telemetry storing:

- Continuously stored data
- Conditionally stored data
- Event log data
- "Keep latest" -data
- Not stored data

Continuous data is the simplest and usually the most voluminous type. A measurement is made at set intervals and stored as the measurements are made. It can be processed before, for example filtered for high-frequency noise. An example of a measurement that would be stored continuously is location. Analyzing the traveled path is useful, but there need to be enough measurement points to rebuild the path. In a case where the results cannot or simply aren’t transmitted live, the location needs to be stored every few seconds in order to recreate the correct path even through the highly branching city streets.

Conditional data is similar to continuous data, but only for fraction of the time the data is stored. The measuring can be constant and the measurements are simply discarded rather than stored when the condition is not met. Conditional data is somewhat hard to define as it is in the middle of continuous data and event data. Continuous data itself is typically actually conditional, since more often than not the equipment has to be turned on in order to make measurements, therefore establishing a condition for measurements. Also to compress data, default measurements such as zero can be removed and filled by the receiver, creating a non-zero -condition. Events are triggered by conditions, so some conditional data could be defined as event data as well. An example of a measurement that would be considered conditional is the amount clutch is pressed while braking. The clutch mechanism measurements would only be stored from the time brake is pressed. This data could be used to compare driving habits of letting the car "roll" more instead of braking, which is more economical and ecological [11].

Event data introduces another type of stored data: logs. Rather than having just long periods of measurements, logs are more precise and constructed. Logs are a set of events that have happened, are usually much more descriptive and contain multiple measurements. Events typically contain timestamps, descriptive names for different events and associated signals. [12] As mentioned previously, events are similar to conditional data, as conditions trigger an event and data to be stored. However, event triggers are usually more complicated and require more processing or multiple signals to activate. For example, a complicated event describing "aggressive driving" would have timestamps for the beginning and the end and contain the behavioral pattern recognition signals such as fast accelerating and braking, honking etc. An event can also be simple, such as "hard braking", which could have only one timestamp (since the event is short) and measurements of just the brake pedal usage. [13]

"Keep latest" -data is extremely simple but important in some fringe cases. Simply, the latest confirmed measurement is stored and written over every time a new confirmed measurement is received. Therefore the actual memory space required is minimal. The odometer mentioned in Chapter 1 is a textbook example of this. The measurement of movement is added to a small overwriting memory block, a counter, the value of which is transmitted by wire into the dashboard of a vehicle. "Keep latest" -data is most common in cases such as these, where the storage is being read continuously and the historical data is irrelevant. Another use for "keep latest" -data is when transmission of data is performed upon request rather than automatically. When transmission of data is only upon request there might be cases where measurements are made upon request as well, but for low-latency communication, the "answer" to the request should be ready before the question is even asked.

Data that is not stored could be overlooked when talking about storing data, but it is relevant when talking about the telemetry pipeline. Not storing data is also an option to consider when deciding about the storage type of measurements. Data that is not stored is either transmitted constantly by wire, like most dashboard signals, or it is simply lost from the period where wireless communication loses or cannot establish a connection. Data that is not required to be perfect or analyzed during a longer period, like live tracking data for public transportation, are usually not stored within the vehicle. Which routes the vehicle has driven can be seen from the planned routes if needed later. And whether or not the vehicle drove a planned route can usually be derived from other available sources of data. The live location is used mainly by passengers, and the passengers do not have a need for old location data.

## 2.3 Transmission

The process of obtaining telemetry data ends with its successful transmission. Before obtaining insight from it there is still the big process of data analysis, but after transmission, the data is no longer susceptible to change. The transmission process requires little customization and can be done with cheap, ready parts. Processing is only needed if the transmitted data is sensitive or a more intricate transmitting network is in use.

Transmission of telemetry data can be made either by wire or wireless. Transmission by wire is more certain, secure and simpler, but is limited by the required physical connection. Wireless transmission’s main benefit is the long range and freedom with the cost of uncertainty. Transmission by wire is also faster, but as data volumes are typically very manageable and wireless transmission can be active often or always, that benefit is almost entirely negated.

Transmission has another good split in use cases, automatic continuous transmission and transmission per request. Continuous transmission typically means that no storage is needed since the data leaves the telemetry pipeline almost immediately. Transmission per request on the other hand can usually communicate to multiple parties if needed and can have more latency in the data processing and storage.

A wired continuous transfer is typically close-range and so ordinary it's sometimes hard to regard as telemetry. Dashboard signals such as velocity are continuously measured from the tires and transmitted by wire to the dashboard for the driver to see. Some of these signals are easily noticed by the driver even without the measured data, for example whether the driver's seat belt is on. These are also the most low-latency and critical signals. While the driver may be able to approximate the vehicle's speed, due to effects like velocitation, commonly known as "highway hypnosis", the display of the objective measurement is invaluable [14].

Wired transmission by request is the classic old vehicle telemetry type. Vehicles perform their area-specific tasks for a period of time, and as they are brought in for maintenance or simply parked for the night, a mechanic plugs in a cord to the vehicle and extracts collected data. The black box in aircraft is the most iconic example of this [8]. While the transmission trends have shifted towards wireless transmission in a way that the vehicles brought for maintenance no longer are plugged in but rather use the hall WiFi instead, it is very common in older vehicles. It also has the benefit of being safe, secure and simple.

Wireless continuous transmission is the rising trend in telemetry transmission. Since internet connection is almost omnipresent in first-world countries, it has also become much more reliable in the last decade. Public transportation such as buses and trains can continuously submit their location to the server [15]. Each possible passenger can then use a smartphone application or a website to track the vehicle without having to question when they will arrive at each stop. Wireless continuous transfer can also be used for less important data, where the data unsuccessfully transmitted is lost rather than stored.

Wireless by request transmission can simply be a wireless version of the traditional black box plug-in at maintenance, but it also has the most potential in more sophisticated systems. In the research on self-driving cars, it has become apparent that wireless communication with nearby vehicles has a grand impact on safety. Of course, there will be a very long transition period, when there will be only some self-driving cars in traffic. Self-driving cars are further discussed in Chapter 5.1 [16].

An important thing to note when discussing both telemetry storage and transmission is that a signal is not always limited to a single purpose. The pipeline may branch into both long-time storing and immediately transmitting data. The branching is not limited to just two, either. Due to the low cost of memory, there is also a very common case of redundancy in stored data, which causes the storage of signals more than once [17]. This is also present in signal derivatives, where distance traveled could be derived from location data. In signal derivatives, usually the accuracy of each measurement is an argument for it, though.

## 2.4 Data reception and processing

When considering fleets of vehicles and multiple vehicles' data, the transmission receiver faces multiple challenges. While transmitting the data is quite simple and transmission and storage solutions can be vehicle-specific, to the receiver the whole picture is more than the sum of its parts. Low-latency data such as location should pass through to the appropriate server as soon as possible, while emptying black boxes causes transmission volumes to spike randomly. More often than not, a flexible solution with cloud technology is selected.

Cloud service providers such as Amazon Web Services, AWS, Microsoft Azure as well as Google Cloud offer services that can handle the shifting characteristics of fleet data. Each of them offers long-time storage for large amounts of data and fast databases for low-latency look-ups and storage. Websites and applications can also be hosted in the cloud, making the data accessing that much simpler. The cloud services scale in a "pay-for-what-you-use" -manner, so shifting volumes of data are not difficult to resource in computing power or money. [18][19][20]

Cloud services also provide processing capability for data. Especially more intensive data processing algorithms would require expensive hardware in order to process something in a reasonable time. The "pay-for-what-you-use" -approach is thus extremely beneficial in processing as well. The interconnected cloud architecture of data storing and processing services also reduces the complexity of moving the data around with ready code libraries. These are the theoretical benefits, at least. In practice, a person must first be quite familiar with the cloud environment before even the simplest of processing jobs can be ran successfully. [21]

Some of the data can be processed before transmission in order to reduce the amount and complexity of data stored. Edge-processing can be as straightforward as summing up measurements to daily totals or averages. [7] The transition from hourly to daily totals for example reduces the data amount to one twenty-fourth. There is however no way to recover the lost complexity, and simplification of data should only be made if nothing of relative value is lost because of it. Processing in the form of data compression does however allow a reduction in transfer amounts with only a processing cost during compression and decompression.

Compression can be applied for both lossless and lossy data volume reduction. In compressing files, data redundancy is harnessed to reduce the size of files into a single file that takes less space than the sum of its parts. Compressed file types such as ZIP, RAR, and CAB can be then transferred and a reverse algorithm applied to obtain the original files. Lossless compression means that the files will be exactly the same after compression and decompression, which is imperative in data such as text. With numerical measurements, lossy compression is more considerable, as it may not matter that the data might be very slightly altered. Also data such as pictures, audio and video can be compressed further than they typically already are without a distinguishable difference to people. The compression and decompression process typically isn't that time-consuming, especially if the data has high similarity. [22]

In Amazon Web Services a model architecture for telemetry data processing could be as seen in Figure 2.1: For low-latency data such as location, an overwriting upload is made to a relative database in DynamoDB, from where an EC2-virtual computer refers to it while running a web service [23] [24]. Larger chunks of data containing the low-latency information as well as other measurements are uploaded to an S3 bucket for longer storage [25]. Larger chunks of data typically require some processing such as decompression, so a Lambda function can be written to process the compressed data as it is uploaded to an S3 bucket [26]. Data processing jobs can be run with Glue for repeated or simple but heavy processing. Sagemaker notebooks or EC2-instances can be used with the clean data for analytics and client applications respectively [27] [28].

## 2.5 Data analysis

After the data is successfully at hand, the next and final step would be to do something with it. Simply storing away massive amounts of data does not provide any business value by itself. Besides the marketing bragging rights to say that the equipment collects a lot of data, at least. A data analyst could apply clustering algorithms to blindly attempt to find patterns in unknown data, but since vehicle telemetry measurements are relatively easy to understand, most of the analysis is hypothesis-driven.

Vehicle telemetry data can be used in hypothesis-driven analysis in several ways. Obviously, it can be used to prove hypotheses such as driving from work is more prone to accidents than driving to work [29]. The extent of proven hypotheses can be measured more accurately, for example exactly how much more dangerous it is to drive tired [30]. The data can also create possible new hypotheses, which require more analysis and possibly entirely new measurements.

Calling data analysis the final step is somewhat misleading, as it affects what should be measured in the first place. Therefore the process of data acquisition and analysis is more circular than linear. Finding possible business angles from existing data will not always

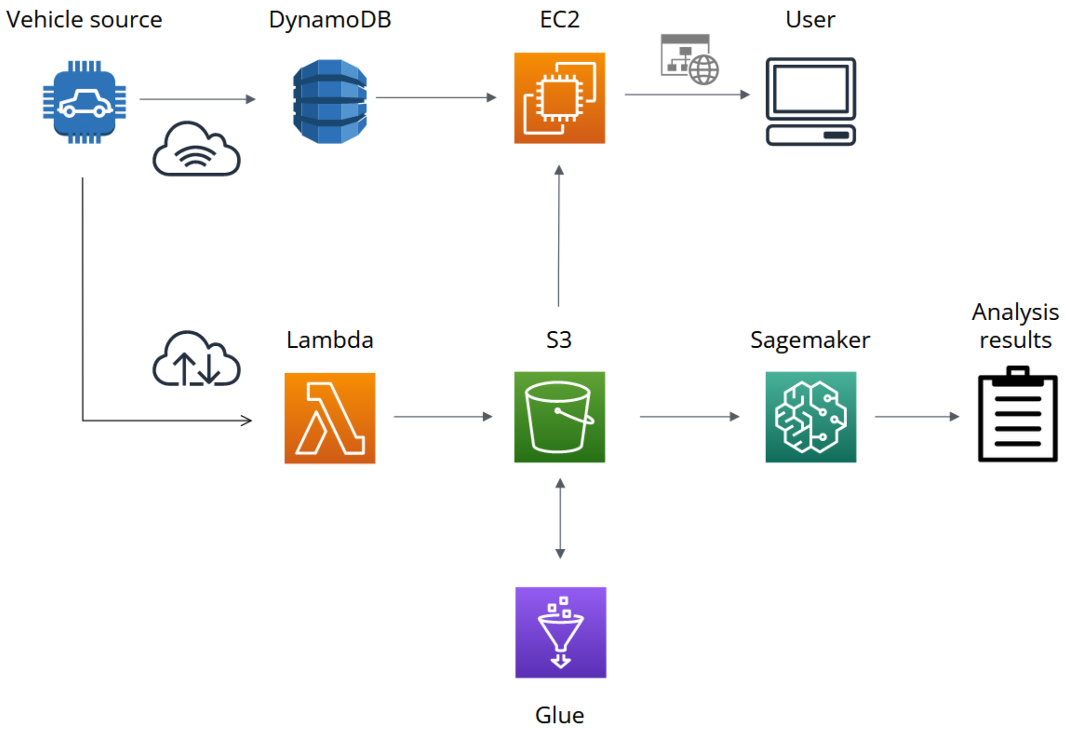

***Figure 2.1.*** *AWS dataflow diagram*

convert to value without more specific or additional measured metrics. If for example the most common cause of trucks breaking down on the road is suspension according to maintenance reports, perhaps the suspension levels could be measured and such costly events prevented. [31, p.4-5]

There are numerous different areas of analysis for telemetry data, but the main topics in this thesis are:

- Correlations
- Trends
- Patterns
- Anomalies
- Optimizations

Correlation is the extent of the linear relationship between two variables. Many physical variables correlate fully, such as force and acceleration when mass is constant. Variables can also be inversely related, like the amount of time a distance takes when traveling with a velocity. With data in the real world, variables rarely correlate perfectly, but finding the correlation coefficient allows accurate estimates. For example, in trucks fuel consumption and cargo weight have a positive correlation: the fuller the truck, the more fuel it will consume on a route. From data, it is easy to find the function between cargo weight and fuel consumption, which can then predict the amount of fuel needed for the route. This prediction is useful in multiple ways and can be used to save money. Knowing the amount of fuel the vehicle will consume allows the calculation of the cost of the trip compared to the benefits, delivery fees. The gas tank of a vehicle also might not have to be fully filled, avoiding the added cost of driving the unnecessary fuel around. [32]

Correlation infamously does not always infer causation. Motorcycles and convertibles are expected to crash into one another more likely relative to the amount of them compared to all vehicles. With this information, one could infer that the drivers of these vehicles are perhaps reckless and drive at higher speeds and are therefore more at risk of crashes. However, as both motorcycles and convertibles are usually driven during similar weather conditions of warm temperatures and no rain, they simply mostly appear in numbers together in traffic. Data analysis can be done by normalizing the amount of each type of vehicle in traffic, after which it is apparent that these two vehicles are not especially dangerous to each other.

The search for causation is also present when analyzing trends. Trends are typically analyzed either in a long time period window or after a distinct presumably causal event. For example, if the average fuel consumption of a company's trucks is on the rise compared to the last few years of data, it might be worthwhile for the company to try to slow down or turn this trend around. After giving incentives for more ecological driving, it would be possible to see from data if the average fuel consumption would start to go down. Trends are also useful in predicting the future on the basis that the present trends continue. [33]

Patterns are more complex and typically more multidimensional than trends. When analyzing long periods of time, a time series of data, similar occurrences that appear in it are called patterns. Patterns can be found in many severe malfunctions of vehicles, so detecting the start of the behavior is imperative in order to stop the malfunction from happening. Very few parts of a vehicle break outright, rather they wear down until they eventually give up. Some of these patterns, such as the violent shaking of a car when a tire starts to go loose, are easily spotted by the driver. Other patterns, like engine wear, are much more intricate, but might still be detectable from telemetry data. [34]

Anomaly detection is also typically learned from and used on time series data. While patterns repeat similarly in data, anomalies can be different each time. From a long time period of learning data, an understanding of the normal fluctuations and values measurements is gained. Patterns are linked to certain repeating and known events, so it allows understanding and reacting to something that is about to happen. As anomalies can be caused by a number of unknown reasons, reacting to them can be difficult. In vehicles, where everything is supposed to run safe and consistent, any large deviation from normal can be assumed to be a bad sign. If the cause for the anomaly is found enough times, it can also be used to train a pattern recognizer to discern these events in the future. [35]

Optimization is the area of telemetry analysis where business knowledge is most important. Almost if not everything can be optimized, and doing it solely with data usually leads to poor results. Optimization is most effective in cases such as route selection. Whether to choose a longer but faster route or a shorter, slower one is a prime example of this. The benefits of saved time are weighed against the extra fuel consumed. A taxi charges the passengers by traveled distance and the time saved might allow for additional trips, the longer but faster route is usually preferred. A chartered bus however is typically paid the same regardless of distance, and is rented for a time period where extra time does not matter, so selecting a slower path is more economical for the business. [36]

## 2.6 Data visualization

Data visualization, often overlooked, is key in both messaging analysis findings as well as finding them. Visualizing telemetry data to people can also be sold as a service, it does not need to happen to just executives behind closed doors. Correlations, trends, patterns and anomalies are much better explained via images than just numbers. Searching for them can also be done by visualizing the data first, and deciding whether trying to fit a line or curve into the data is a worthwhile effort.

Data can be visualized in multiple ways. To most common are two-dimensional plots, where the x-axis is time. Both the developments of metrics over time and recurring patterns in shorter periods of time are easy to understand and informative. Plots can also show multiple different lines in one image, without overwhelming the viewer. There are visualizations that work for some data exceptionally well, like maps for location and charts for percentages. Most two-dimensional graphs cannot however visualize multidimensional data. A third dimension can be arbitrarily added by color or hue, but that is of little use when the number of dimensions is in the dozens. [37]

Real-life metrics usually have correlations to more than one metric, making two-dimensional visualizations difficult. The most obvious way to plot relationships between multiple metrics is to create a two-dimensional plot for each pair of metrics. The number of plots however gets out of hand quite fast when the number of dimensions grows. Much like the third dimension can be added to a plot by introducing hue, the fourth, fifth and sixth dimensions can be respectively added with depth, size and shape. The interpretative accuracy does begin to decrease this way, as seen in Figure 2.2 [37]

A picture is worth a thousand words and as such the explanation of the contents of data is best explained visually. Especially if data is analyzed by a professional, the findings can be simplified to a level any likely recipient is able to understand. The ease of understanding is not measured only in the complexity of explanation but in the time it requires to process as well. Just a single glance required to understand a well-constructed graph might be the value a person brings to a company, saving others' time. [38]

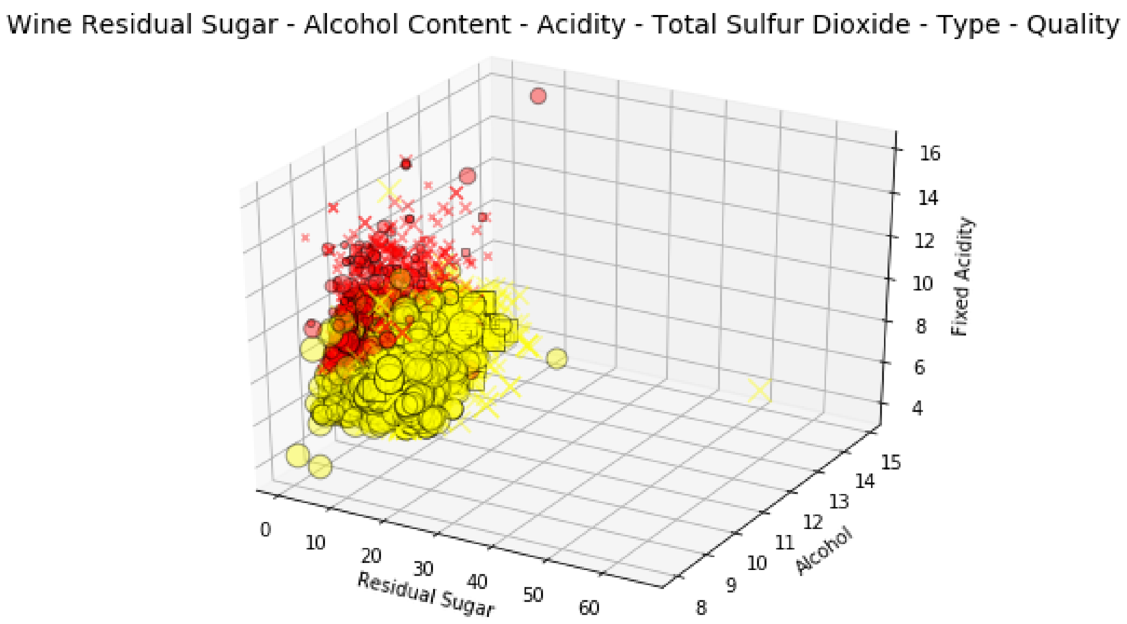

*__Figure 2.2.__ A six-dimensional plot of wine properties [37]*

A common saying states the three categories of lies: Lies, damned lies and statistics. As graphs are visual representations of larger amounts of data, the definition of a statistic, there are both pitfalls and responsibilities tied to it. Attempting to forcefully fit curves or patterns into a data might eventually make a promising hypothesis, but these hypotheses rarely hold when including additional data. The responsibility linked to this is to create visualizations so that they do not convey false assumptions. Correct axes for reference, the entirety of data, and enough relevant information about its acquisition should be included in each provided graph. [38]

# 3. VEHICLE TELEMETRY

In this chapter, the scope of telemetry is narrowed down to vehicles. A brief look back in history is made to better understand the direction and speed of development in vehicle telemetry. Operation critical telemetry signals are introduced first as they affect all vehicles. Afterward, the focus is set to delivery and transportation. The metrics in these fields require more complicated data analysis and tie in closely to applications discussed in the next chapter.

This thesis limits the definition of a vehicle to select ground vehicles. More accurately, vehicles that operate on rails and motor vehicles of EU vehicle categories M and N. Essentially this grouping includes trains, trams, cars, vans, trucks and buses. These vehicle classes have a lot of common themes and the evolution of telemetry use can be shown with them. [39]

## 3.1 Railway telemetry

### 3.1.1 History

Roads existed long before railways did, but from vehicle telemetry's point of view, railbound vehicles started the gathering and reporting of data related to operating vehicles first. While the history of rails is significantly longer, steam-powered locomotives started appearing on rails at the beginning of the 19th century. Back then, there were both double- and single-tracked roads, and wheels came off tracks easily. As most railroads were a monopoly, during a derailing one would send a messenger horseback to the nearest outpost and get a crew to re-rail it. Any other trains would have to wait behind the derailed train. However, as there was a very small number of trains, this was not as big of a problem. [40]

The invention of the telegraph revolutionized railway operations. There were earlier attempts to convey signals along the railroad such as pneumatic whistles, but the telegraph was cheaper and more reliable. The telegraph lines installed in Britain were so successful that over 15 000 remained active at the turn of the century. While the train itself was still unable to send any signals, the outposts could communicate in real-time. Outposts would report passing trains, resulting in the first transportation tracking service. And if the train would not arrive at the next outpost within the expected time, a re-railing crew could be sent to search for it along the railway, significantly reducing the amount of time these accidents took. [41, p.66-69]

For the following hundred years, very little would change in railway telemetry. The trains would continue to get faster, more reliable, quieter and their number would increase, but they would remain fundamentally the same. A passenger would go to the station, buy a ticket and be able to find out the predicted time the train would pass by. The telemetry inside the train did advance, as telemetry such as speedometers and engine temperature became available to the conductor. The trains could communicate with stations via radio between outposts, but that wouldn't be visible to the passenger.

### 3.1.2 Modern railways

Digitalization and the internet define the modern train experience. The live location of the train can be followed, its current number of carriages and even the dining car menu are available to the public in real-time. Tickets can be bought and seats reserved. Even the other reserved seats are visible to the customer, allowing him to choose between sitting alone or in the company of strangers. In Nordic countries, it isn't even possible to buy train tickets physically anymore. Inside the train, everything that a passenger could possibly want to see from current speed to inside temperature have been brought to monitors. [42]

Traditional trains aren't the only vehicles that operate on rails. Modern metropolises are defined by their underground railways, where subway trains pass by every few minutes. The extremely low travel times and constant travel speeds between stations render live location quite irrelevant to the people waiting, as the estimated arrival times can just be updated. However, from the inside, the view is just dark rock walls, and current location cannot be deduced from them. Location displays can operate simply by preset timers, as the travel times between stations vary only by seconds.

Where larger trains would not fit on the city streets and subway routes would be too cost ineffective to build, there can still be smaller railway trams. Trams are a more ecological alternative to buses, with the downside of their routes literally being set in stone. As trams operate inside the street traffic, they are a lot more susceptible to delays. Because of this, trams are typically equipped with wireless transmitters that broadcast their current location, so timetables and smartphone apps have the exact estimates for each tram stop [43].

## 3.2 Automobile telemetry

### 3.2.1 History

A century after steam locomotives, the first automobiles started appearing. Horse carriages still dominated streets for a long time, as cars were a luxurious commodity. The cost eventually dropped down and middle-class citizens started to afford and obtain motored vehicles. The first cars had pedals, steering wheels and a winch to turn the car on, but besides that, no information was passed to the driver. At the time the models available were as slow as horse carriages and more so if carrying extra weight, so there was little need for them. [44]

With more cars on the streets and the cars getting larger and faster, road safety became a large concern. When the first cars would come to a collision or a sudden stop, the passengers would fly off the car, resulting in some cuts and bruises or the occasional broken bone. However, as cars were beginning to move over 60km/h, sudden stops in old cars were more often fatal. To prevent the needless loss of life, governments and car manufacturers attacked the problem on multiple fronts.

Laws prohibiting reckless driving were introduced all over the world in the 1920s. Speeding, not slowing for pedestrian crossings and quite importantly drunk driving were all condemned. Voluntary driver education courses were taught in high schools and universities. Later on, driver's licenses became widespread after the United Nations hosted the Geneva Convention on Road Safety in 1949 [45]. To this day, the education required for a driver's license differs by nation, but in modern countries is mandatory. The public opinion was that accidents were caused by driver inadequacy, but the car manufacturers soon realized that some accidents were unavoidable and thus began the inventions of several iconic safety features. [46]

Cars have two categories of safety systems, which began to develop. Active safety features help the driver to avoid an accident altogether, while passive safety systems mitigate the damage of an accident. Better brakes and grip in wheels for example increase the reaction period in which breaking would avoid or lessen an accident. Headlights were split into near and far lights as well as fog lights. Lights were also used to signal other drivers with turn lights and brake lights being set up in new cars. The drivers would also have access to more telemetry data as sliding warnings, freezing temperature warnings and other signals were added to the dashboards.

Contrary to active safety systems, inventions and innovations increased the safety of drivers and passengers in leaps. Seat belts were invented to hold passengers in place during accidents, which proved to be extremely pivotal in reducing traffic deaths. The standard three-point seat-belt proved to be so effective that as the inventor, Volvo made it patent-free to use in the interest of public safety [47]. Windshields were changed to laminated glass, which stopped injuries from showers of shards of glass upon shattering. The third major invention was the airbag. Fitted on the dashboard as well as the wheel, the exploding pillow would cushion the impact from the accident.

A modern car has dozens of safety systems, some of which the drivers themselves might not even be aware of. The vast amount of warning lights on the dashboard help pinpoint faults that could cause issues during driving. Brakes have automatic anti-lock systems that ensure maximum braking power on slippery surfaces. The spread of force from an impact is directed by the chassis and bumper in a way that passenger harm is minimized. Some of the modern safety innovations might not even be just for high-speed collisions. Automated parking for example works at low speeds, and reduces the risk of damaging other vehicles.

### 3.2.2 Automobile telemetry signals

Cars, vans, trucks and buses all have the same base structure. Chassis, engine, transmission and body may vary in size and shape, but their functionality remains mostly the same. Most importantly, these vehicles are driven very similarly. While larger vehicles require additional licenses to drive, a steering wheel and pedals are used to control each of the vehicles. As such, most of the telemetry available in the basic parts either is or can be collected in each vehicle.

The first category of signals is the immediate signals required to safely and efficiently drive a vehicle. These signals are immediately transferred to the dashboard and typically forgotten immediately as well. These signals include important gauges and some of the more common dashboard symbols. The number of gauges on a dashboard and their appearance vary by model of the vehicle, but typically at least contain a speedometer for velocity, a tachometer for RPM, a fuel gauge and a thermometer for engine temperature. Of these, only the fuel gauge is not immediately necessary for driving. Dashboard symbols, typically red or yellow in color, such as warnings for airbags, engine temperature and handbrake should be reacted to as soon as possible. A speedometer, tachometer and signal light descriptions are listed in Figure 3.1. [48]

The second category of signals is the less important signals, which are still conveyed to the driver. Most of the signals in Figure 3.1 fit this category. Everything from confirmatory signals like "headlights are on" to the "check engine" -light is not immediately critical to driving. However, there is hardly any reason to doubt whether they should be shown to the driver. Odometer reading and outside temperature are usually shown on the dashboard and also fit well into this category. [49]

In modern vehicles, there is usually another screen beside the dashboard mostly reserved

***Figure 3.1.*** *List of dashboard symbols and their definitions [48]*

for radio and such. A wide and varying range of settings can be accessed through the user interface created by the screen and nearby buttons. Besides the sound system, important features such as ABS-brakes and power steering can be toggled on and off. Collecting data from these sources offers a lot of possible value, as the vehicle manufacturer should be very interested if a lion's share of drivers chooses to opt-out of ABS-brakes or power steering. Even something like a significant amount people using bass boost to listen to music might imply there is something out of order in the sound system. [50] [51]

The last category is the signals collected by the motor but not immediately displayed to the driver. A majority of these signals are derivatives of the first and second category signals. The "check engine" -light for example is decided by a number of inter-engine measurements, from which the signal value is derived [49]. Some signals are conveyed to the driver in a subjective way. For example, while the driver can both set the volume of the radio and listen to the amplitude of the sound, it is very difficult to guess the actual decibel value played by the sound system. As there is a varying amount of background noise while driving and humans naturally try to filter it out, the radio volume can rise up to even dangerous levels without the driver actually noticing. Also, the relation between the selected volume number and the actual decibel value of the sound does not fully correlate with for example third-party gadgets. Playing music from phones or MP3-devices with the same decibel values requires different volume numbers due to messaging protocols and varying volume baselines. [52]

It is important to note that the signals from a motor vehicle are transferred to more people than just the driver. For example, toggling a turn signal on does not just display an arrow on the dashboard and a clicking noise, but activates the lights behind and in front of the vehicle. This is also true for braking, as notifying the actions the vehicle is or will be taking is imperative for traffic to run smoothly.

## 3.3 Passenger car telemetry

In Europe, approximately eight out of nine road vehicles in use were classified as passenger cars in 2019 [53]. While buses, trains and such fit more people in them, the vast majority of commuting is done by passenger cars as well, carpooling included. Since passenger cars are so dominant in traffic, the focus is usually set on them in traffic design. The financial benefits of innovations in passenger cars are also much larger than in other vehicle groups.

As most cars are passenger cars, naturally they are also the most bought and sold type of vehicle. The market share for car manufacturers is extremely heavily contested, resulting in fierce competition for the best models in every category imaginable. Whether the most desired attribute in question is high acceleration, low emissions, attractive outlook, vast cargo space or user-friendliness, car manufacturers usually have multiple models with a focus on a few attributes each.

Cars are usually traded multiple times in their lifespan. For reference, in the UK during Q2 of 2019, 661 795 new automotive vehicles were sold [54]. In the same period of time, there were 2 034 236 used car transactions [55]. While the number of car transactions is slightly skewed due to car ownership often moving within households and immediate families, second-hand car purchases are still much more common than those of brand new cars. Therefore approximation of the condition of a used car is left to regular car owners. Cars are not a cheap commodity, so the condition of the vehicle is the most determining factor on how many thousands is the car still worth.

The condition of a car is most accurately measured by a skilled mechanic. However, the services of one do not come free, and especially with older and cheaper cars would take a sizeable margin of the cost of the vehicle. Some of the measurements a mechanic would take, from battery leakage to tire wear, can also be performed by regular individuals. However, as a large portion of people owns a car, the equipment and knowledge required to make these measurements typically prove too high a hurdle. The average car purchaser is interested in the simplest and most expressive metrics as well as a test drive [56].

The total driven distance is the most common metric discussed during a second-hand car sale. As mentioned in Chapter 3.1, the odometer measurement is typically found on the dashboard for easy access for anyone to check. As a car's primary function is to be driven around, the amount of driving done is reflected in the odometer reading. A car has an expected lifespan of some hundred thousand kilometers, the value could be expected to be linear to the amount of expected lifespan left. A car expected to travel 200 000km would after 100 000km then be worth half of the original. [57]

The wear of the car is unfortunately not linear. As towards the end of the car's lifespan, its parts deteriorate one at a time. For example, if the battery dies, changing it could maybe allow another few thousand kilometers to be driven before something else breaks. These repairs add an additional cost to the car owner and should reflect the price of a second-hand car. Some of these repairs can also simply not be made due to cost-effectiveness. A minor fault, for example seat warmers not working, is hardly worth the trouble or the money to fix them if the owner expects to drive the car only a couple more months during the summer. These minor or major flaws lower the value of a car even less, as they make it feel worse to drive.[57]

The driving feel is at least as common as the mileage when assessing the condition of the car. Even with new cars, a test drive is expected, as the majority of flaws do not show to the outside. A car might simply just not fit a driver's tastes either. There are many features that different drivers prefer from the number of gears to the existence of the clutch itself if the car has automatic transmission. Or even really small features, like the stiffness of the wheel or the clutch, which be deal-breakers for people that cannot adjust to them.

While the driving feel is entirely subjective, all of the different aspects mentioned above are measurable. Battery leakage can be measured by recording its voltage over a time period where it is not charged or used. The transmission type and the number of gears are most definitely in the user manual, but the stiffness of the wheel and the clutch are not. The stiffness can be measured by the force required to move them in newtons. Even seat warmer functionality can be measured by the amount of power the warming drains from the battery. However, there are two inherent problems in the approach to express all of these in data. First, telling that the clutch takes exactly ninety-five newtons to push down has no informative value to an average individual without a point of reference. Second, the measuring requires equipment or time, which again are not free. [58]

The opaque measurement results could be made simpler by having a point of reference. If a potential buyer has the same measurements from a car he has driven before, he can make easy comparisons between them. A five percent decrease in the force required to turn the wheel should be barely noticeable, but someone with weaker knees should probably stray from a car with double the force required to operate the clutch. However, this approach to expressing different aspects of the driving feel with data amplifies the second problem. The value of measuring these aspects would require not only the car to be bought to be measured, but previous cars as well. [58]

The driving feel is not applicable only to old cars and their faults. While faults deduct from the driving feel, helpful features and ease of use add to it. The benefit of installing new positive features is also more immediate and easier to sell than spotting faults in the future. Newer models of cars fiercely compete with their latest innovations and ways to bring their costs down. Driving has been made easier with features like constant speed drive and cruise lane control. While relinquishing control to automated systems is not desirable for some drivers, there are also features that give more data to control the vehicle, such as reverse cameras. New car models have an extensive user interface on a monitor on the dashboard for everything from driving data to air condition settings. [59]

With essentially a fully capable laptop on the dashboard, less important data has a way to be shown to the drivers without overwhelming them. This less important data can be split into three groups:

- Data that can be explained further
- Data that would not fit the dashboard otherwise
- Data that is only of interest to a subgroup of drivers

As discussed in Chapter 3.2.2, some dashboard lights are derivatives on inter-engine signals. Engine warnings or typically the ’check engine’ -light are derivatives of many signals that signal faults in different parts of the engine [49]. With a monitor and an interface, the exact reason why the warning is active can be shown. If a battery warning is active for example, the interface can explain whether it is because the voltage has dropped too low, the charging has troubles or the voltage spikes due to connection issues. This access to telemetry may allow the user to either fix or assess the need for repairs without consulting a mechanic. [50]

There is limited space on the dashboard for signal lights and meters. While the example above about signals consisting of multiple other signals could be seen as an example of reducing the required dashboard space, there are also signals that can only be presented well in a manner that needs space. All of the signal lights are binary, which limits their ability to represent data. This is why metrics such as vehicle speed and engine rounds per minute are presented as analog meters. In order to be readable, the meters have to be quite large and as such, only the most important ones are available in older cars. However, with the option of an interface for miscellaneous signals, a lot of slightly less important measurements can be shown on request. The amount of other consumables besides gas, such as oil and washer fluid, are just as measurable as gas is. Similarly, the air pressure in tires is not likely to change that fast but traditionally would require the driver to measure them physically.

The last group of data measurements that can be shown is not necessarily shown due to their objective usefulness but because of their ranging desirability. Some people consider the ecology of their driving to be the top priority, while others do not care about it at all. As such, it is not good marketing to feature the amount of fuel burnt for example powering the air conditioning or playing music on the dashboard for everyone. On the interface, however, these things can be found by people who look for them. A beginner driver is likely to be curious about his bad driving habits that can be spotted by telemetry, while a veteran driver will very likely not want his driving skills to be questioned. [51]

The interface is not limited to just data representation, as it can replace some of the older dashboard controls as well. Car radio can be relocated to the interface and doesn't need to take space on the dashboard. Music volume, air conditioning and parental controls can easily be accessed and can be left on top of the interface during driving. Ultimately, the interface can reduce the amount of occupied space on the dashboard to provide a less cluttered look. Some car models, especially ones marketed to families might even have simple video games available to be played using the interface, in case a person in the car requires entertainment while the car is idle. [51]

## 3.4 Delivery telemetry

Vehicles move their driver for long distances fast, but they are by no means limited to that. Most vehicles fit passengers in addition to the driver and have room for cargo as well. Passenger cars have a mix of both, but the focus is usually on one or the other. Buses are designed to transport the maximum amount of passengers while a semi-trailer truck fits only two people for tens of tonnes of cargo. Railroad cars are similarly either passenger wagons or cargo wagons, which can be linked together to fit transportation requirements.

Transportation requirements also vary in a very hierarchical manner. To optimize costs, Larger vehicles are used for larger amounts of goods and smaller vehicles are used to deliver smaller loads to final locations. Ordering a package from the opposite end of the world is first brought to a post office via a small car or a van. The package is then bundled with other cargo heading in the same direction, and a larger vehicle is loaded with them. The larger van or a truck then delivers the package among others to the next post office and eventually to a port. From the port, a massive cargo ship then completes the journey to the other side of the earth. [60]

As the massive cargo ship arrives in a major port, the delivery vehicles get smaller between each stop. A smaller cargo ship might move the package to a minor port, but heading inland the package is loaded on either a large truck or a train. These make the opposite journey from larger delivery centers to smaller ones, and smaller vehicles then continue the delivery to the post office, from which a car or a small van finally delivers the package to the recipient's doorstep. [60]

The optimal size of the vehicle required in each step can easily be asserted. Naturally, the maximum cargo space and overall vehicle size correlate highly. Also, the larger the vehicle, the heavier it is. The amount of fuel consumed is based on the sum of the weight of the vehicle and the cargo it carries. Thus, a large vehicle carrying the same amount of cargo as a small vehicle consumes more fuel. Larger vehicles are also more expensive and their wear is more related to the distance driven than the cargo carried. The optimal size of the vehicle is thus the smallest vehicle possible that can carry the required load. For a large delivery company, the reality is not as simple. The transport amounts vary, and there are opportunities to either use a large vehicle or multiple smaller ones.

While the delivery vehicle should be as large as possible, there are upper limits to this. A width limit is apparent on roads, as the vehicle isn't allowed to extend to other lanes. A physical limit in length is less apparent, but a longer vehicle is a lot clumsier in navigating turns or on uneven surfaces. The last axis, height, is restricted by overhead bridges and tunnels. These dimensions slightly vary by country, but for example, in Finnish law, the largest allowed vehicle would be a trailer truck 34.5 meters long, 2.6 meters wide and 4.4 meters high. Another factor limiting mostly height is the weight of the vehicle. Roads will considerably wear if exposed to too much pressure. The pressure of the vehicle's weight can be lowered by adding more wheels, so the weight limits are typically tied to amount of axles a vehicle has. The aforementioned maximum size truck could weigh up to 76 tonnes if it had eleven axles. [61]

As the weight of the vehicle can change depending on the amount of cargo, it is an important measurable metric. It isn't something the vehicle itself has to measure, however. Weight is easily calculated by the sum of its parts. Adding the vehicle, fuel and cargo weight together provides the total weight with enough accuracy. The distribution of weight requires slightly more advanced math, but typically it is desirable to aim for even distribution across cargo space. Even distribution guarantees the most predictable handling for the driver as well as even axle pressure.

As previously mentioned, the maximum weight is related to the number of axles. The maximum weight of a truck is based on the maximum pressure from an axle to the ground. The weight has to be split across axles in a way that no axle surpasses the maximum pressure. If an axle has over the allowed weight, the vehicle is subject to considerable fines. The maximum axle pressure is traditionally measured in weighing stations on country or state borders. Modern weighing systems are weight-in-motion, WIM, allowing the measurement to be made without vehicles having to stop. These systems have mostly replaced the old weighing stations and are also used on some bridges. [62]

The last thing to consider when loading cargo onto the truck is how will it be unloaded. If similar to the first locomotives hauling coal and having only one destination and one type of cargo, unloading is simple. An oil truck can easily make multiple stops and pump the required amount at each destination. However, there often are cases where a truck has multiple types of cargo that needs to be delivered to different destinations. In these cases, it is extremely unwise to load the first destination's cargo to the back of the cargo space, as everything else would have to be unloaded to access it on the first stop. With maximum weight on axles, maximum physical dimensions and a functional unloading order, loading a delivery truck optimally is a difficult multivariate game of Tetris. [63]

Closely related to unloading, the driver and manager of the vehicle should always be aware of what the vehicle is carrying. At its least, which cargo is unloaded at which stop has to be known. If the cargo is prone to breaking, toxic or even explosive, measures to drive safely have to be taken to avoid accidents. It is also very possible that an error is made in the delivery center, or a product recall is called while they are still en route. Being able to track down the correct vehicle before cargo is unloaded at the wrong place can save a lot of money, as no additional vehicle has to be sent to move the cargo to its correct destination. Cargo is typically specified in a cargo manifest, with a physical copy on board the vehicle itself. [63]

One big helping factor in delivery vehicle loading is that delivery is often repetitive. Most businesses require a constant supply chain, such as gas stations and post offices. Even grocery stores, which can order ever-changing sets of goods, still require a delivery made daily. Therefore, most of the loading choices and delivery routes do not constantly need to be remade. By not having to make large adjustments to routing, the option to carefully optimize them is also possible. At best, a vehicle is on the move around the clock with as much cargo as possible. This does require switching drivers on the fly, however, as drivers require sleep. The routing then has to account for pickup and drop-off points for drivers, who have to be able to conveniently travel to those points.

## 3.5 Transportation telemetry

### 3.5.1 Problem setting

While the transportation of goods might be greater than the transportation of people in mass, the majority of vehicles are used to transport people. More specifically, designed to transport people, as most vehicles are also able to carry cargo to varying extents. The amount of people transported in a vehicle also vary, resulting in a wide array of possible methods of transportation. For example, much like with deliveries, transportation too has inherent hierarchical properties. Car ferries are driven onto in small groups, the ferry ride is taken together, and afterward each car goes their separate way.

People aren't cargo and shouldn't be considered as such, but there are similar qualities. Besides the aforementioned hierarchy in long-distance travel, another key similarity is the repetitiveness of transportation. As mentioned in the previous chapter, repetitiveness is a very helpful factor in the planning of transportation. The average person wakes up in the morning, goes to work and returns in the afternoon. The person typically also has recurring activities such as hobbies and shopping which happen in fixed locations. The place of employment or education will most likely change during the years, but they rarely remain constant for less than months at a time. As such, if the person requires transportation with other people like him, the routes and timetables could, and in reality are, condensed to those parameters. [64]

A clear contradiction to cargo transportation, the net sum of the movement of people is very close to zero. For delivery, transportation is made and the cargo delivered is left there. Returning empty is very inefficient, so entirely different cargo is loaded and delivered somewhere else. The infamous triangle trade across the Atlantic is a key example of this. People, however, will want to return home at the end of the day. Because of this, back-and-forth transportation has a lot of demand, and public transportation routes are almost always linear. Even when considering private motoring, it is expected that the vehicle will eventually return to where it set out from. [64]

Public transportation and private motoring should always be considered separately, even if they do share the roads. The relationship between them is not always intuitive and there are multiple cascading reasons that result in people opting for either one. Actions such as adding more public transportation for high traffic areas could be considered smart, as people are more tightly packed in public transportation and the maximum throughput of the road would increase. However, if enough of the people are simply heading further than the buses or the buses are felt to be uncomfortable to travel in, the high traffic area could see an increase in traffic, as the public transports would be sparser inside. [64]

With a lot of uncertainty and speculation in cause and effect, only large overhauling actions or trial-and-error -approaches can be effectively taken. Neither of these is a desirable option and the combination of both is even worse. Decisions shouldn't be made without wider knowledge of the problem itself ,as it will reduce the decision to a gamble. Using data as a decision-making base adds credibility to proposed ideas. To use data, however, it must be gathered first.

### 3.5.2 Data use in transportation

The best kind of data to use in public transportation is public opinion. If speculating why people opt in or out of public transportation, the simplest solution is to ask them for the answer rather than just guess. Receiving answers from enough people will allow statistical and clustering methods to be used to better estimate which results would possible actions yield. Having customers' opinions listened to is also an added benefit and generally increases their opinion of the business. A small-scale survey is also extremely easy to do, as it only requires somehow broadcasting a method of feedback to those that wish to voice their opinions, for example an e-mail address listed on a website. [65]

There are several difficulties in customer surveys as well. While just mentioned that some people will go to lengths to give feedback, the vast majority of people will not. As such, gathering voluminous data requires more effort or money. The amount of gained feedback can be increased by making the option to give feedback as easy as possible. Popular ways for this include notifications of the possibility to give feedback via e-mail or as a notification in a smartphone application. Money can also be used as an incentive, either separately or in combination with the aforementioned methods of accessibility for giving feedback. Most common approaches are either direct payments in form of price reductions for the next purchase, or raffles that are entered by giving feedback. [65]

With enough data gathered from the public, some troublesome features can be found. First of all, a lot of people's opinions are still missing, as it is practically impossible to have everybody answer. The received answers are from a subgroup of all customers, which can be aptly named as "people who answered the survey". This group does not necessarily represent the opinions of the entire customer base accurately. Second, there is a lot of noise in the data. Psychologically, the mood of people changes perpetually, and the answers given reflect only the mood they were in [66]. More universally, people and their opinions vary in a very wide range, resulting in a lot of different points of view. Lastly, there might be a negative impact on the approach selected to gather the data. If the application notifications or e-mails are too frequent they might annoy people, or if the monetary gain of giving feedback is too high, there might be a significant amount of surveys filled as fast as possible without any thought given.

The answers received from the public are completely subjective. Different people traveling the exact same way may have completely opposite experiences when asked about it. One person might consider the transportation to be behind schedule, while another person thanks the vehicle for waiting for passengers that arrive on stops in the nick of time. A ride could be considered to be loud, shaky and uncomfortable, while others find it easy to fall asleep during the journey. As such, for every complaint about an aspect of travel, the opposite should also be considered. This is important due to the fact that submitted reviews are more likely to be motivated by negative than positive experiences [67]. Attempting to fix something for a small outspoken group of people might reduce the average opinion and cause opposite reviews to appear en masse.

As the need for objective data has become apparent, the largest hurdle is to decide what kind of data to gather. The customer surveys could be used as a guide, but the planning and installation of data gathering equipment are too time-consuming to be used as a reactive answer. Especially as the amount of transportation vehicles grows, the customer needs should rather be anticipated than reacted to. Brand new inventions might not have comparison points, but for a lot of possibilities, references can be found globally.

### 3.5.3 Transportation metrics

In their 2021 report, McKinsey & Company ranked transportation systems in 25 cities globally [68]. Some of the transportation systems in the report, such as shared electric bikes and air travel, are out of scope for this thesis, but the ranking aspects for conventional public transport are still extensive. The approach in the ranking is objective measurements that should reflect the condition of the transportation systems. The metrics in the report discussed below are grouped as follows:

- External factors (Route network, ticket cost)
- Efficiency (Average speed, amount of waiting on swaps)
- Safety (Average vehicle age, followed safety standards)
- Ecology (Fuel consumption, fullness)
- Convenience (Electronic services, accessibility)

External factors are from hard to impossible for a transportation company to have an effect on but should be noted nonetheless. The company has to operate in these conditions and should make the best possible decisions within the given confinements. For example, a city's route network might have the shortest routes through the city be extremely slow due to the vast amount of traffic. A company that acknowledges this can either try to move as many transportation routes to circling paths as possible or lobby the city representatives to limit private transportation within these roads by allocating lanes to public transportation only. It is important to note that the cost of a ticket for public transportation has external factors that affect it. These factors can be for example the price of fuel and its taxation in the country or state or city support for public transportation, which can artificially lower the cost of a ticket.

Efficiency in [68] describes the speed a passenger can move from a random point A to a random point B in the city using as much public transportation as possible. This is different from the business version of efficiency, where the number of passenger fees are compared to the number of vehicles, routes and the amount of fuel consumed [69]. The speed in this case is measured by the time spent over the straight distance from A to B, as moving fast in suboptimal directions does not benefit the passenger. Not all of the distance can always be covered by vehicles, meaning that the hypothetical person will swap vehicles and travel the rest by foot. Walking speed is the minimum speed, which will occasionally be the fastest method of moving. This typically happens in very short distances, but can also happen along longer paths, if the traffic is dense, swaps would require a lot of waiting or there are no alternative methods of travel available. Efficiency is an important metric that can be improved by acknowledging the main paths passengers want to take as well as spacing connecting routes efficiently when straight paths cannot be utilized.

Safety is a combination of multiple metrics, as generally, public transportation is extremely safe. Fatalities do happen, although rarely, as even modern trains and subways can derail and buses can crash. However, as these occurrences are extremely rare, typically singular, the entire transportation system cannot be penalized by an outlier. There are types of safety that can be measured, though. Minor accidents reduce the feeling of safety for passengers, and transportation culture may leave especially young women vulnerable to sexual assaults. To measure the likelihood that a vehicle will perform as it should, which should be by default safe, the average age of the vehicle and the frequency of safety inspections are good indicators. There also are different safety standards, the effectiveness of which can be approximated and at least somewhat ranked. [70] [71]

Ecology is a simple concept but can affect many aspects of travel. The total carbon emissions for a vehicle can be approximated from data that is easy to find, mainly the beginning and end of lifespan and fuel consumed. Only looking at that as an indicator for ecology is misleading, however. When considering a transportation system, such as a bus network, the average carbon emissions of vehicles by their expected lives must also be divided by the number of passenger miles provided. Simply put, a full vehicle is much more ecological in the purpose it serves than an almost empty one. Of course, opting for more eco-friendly vehicles with lover fuel consumption and longer lifespans is also a way to increase the ecology ranking. Sparsing timetables to achieve better fullness would also provide a similar increase, but packing people to western buses similar to Indian trains would drastically reduce the convenience of travel. [72]

Convenience is the vastest metric and consists of a great number of possibilities with new ones added constantly. The shift from functionality to convenience has been present for decades now, as the vehicles themselves became more affordable. With such a low hurdle to enter the competition, the business advantages must be gained by having customers prefer one service over another based on comfort. The best modern example of this ecosystem is the social driving app Uber [73]. Every person with a car can decide at a moment's notice to start driving what is essentially a taxi. With the market open to everyone, the drivers can attempt to gain an advantage with favorable reviews by providing conveniences to passengers.

Sources of convenience can be split into two parts: Sources inside the vehicle and more general, outside sources. Convenience inside the vehicle stems from every small thing imaginable. Accessibility, ideal inside temperature and comfortable seats are just a few examples of sources of convenience that add up to the whole experience. Different areas of transportation have unique factors, such as the quality of food in the restaurant car of a train. While the last one is of course not measured by [68], it can have a large impact on revenue earned. It is also important to note that many of these inside sources of convenience affect the driver of the vehicle as well and can affect job satisfaction and company loyalty a significant amount.

Outside sources of convenience include the consistency of travel as well as outside services. A form of transportation is more convenient if it is predictable and available. If public transportation can stay true to its schedule, it makes planning and mid-route swaps easier and thus more comfortable. Similarly, if the route is traveled more often, the vehicles are generally less full and the passengers will have to deal with less waiting at their destination. Availability can include things such as the distance to the nearest stop or the amount of time it takes to call a taxi. Modern outside services for transportation are usually electronic, referring to things such as websites and smartphone applications. Electric ride orderings or ticket purchases also reduce the amount of manual labor tied to the traditional methods. Live tracking of vehicles provides hard proof that the vehicles are on their way and route planner applications ease the travel for tourists and other new customers. [15] [74]

### 3.5.4 Use of telemetry

In order to deal with incomplete sampling and subjective answers from customer feedback, objective reference is required. This is the problem that vehicle telemetry fixes. If a sizeable amount of feedback is complaints about timetables being unreliable for a route, the truth can be found by checking the logs of arrival at each stop. If the arrivals indeed are not consistent or are consistently wrong, the problem is real and there is a clear incentive to do something about it. Alternatively, if the timetables are exactly followed, perhaps the physical timetables at stops are simply outdated in a case of a simple mistake. Either way, a course of action is justified and more certain to provide the desired outcome if the facts behind it are known.

The ranking metric grouping introduced in [68] can also be used to group attributes that can be measured using telemetry. For the metrics efficiency, safety, ecology, convenience and external factors, telemetry can be used efficiently in all but the last one. Even for external factors, as mentioned previously, data can be used to either test different solutions to routing or as a justification when lobbying an appropriate government body.

Efficiency is the calculated average speed a person can move using public transportation between random points in the city. Calculating the theoretical average time is an algorithmic problem, where a path-finding algorithm can deduce the quickest route from a large number of possibilities. The planned timetables for transportation can be used with approximated average walking speed to solve the quickest path without the need for vehicle telemetry at all. However, as a famous quote associated to Albert Einstein goes: "In theory, theory and practice are the same. In practice, they are not". The average speed can only be accurately calculated by sampling actualized routes rather than the ideal ones. Unlike theoretical ones, actualized timetables have an amount of variance in them. [75]

While the variance in theory would have no effect on the average speed, the practice is again different. A real person would make the path selection based on the theoretical timetables, usually trying to minimize waiting for a vehicle. A simple example path would be a five-minute walk to a bus station, where the person would ideally wait a minute and then take that bus exactly to the destination. If the bus route in question averages a vehicle every fifteen minutes with a standard deviation of one minute, the passengers realized waiting time does not average one minute. If the bus is over one minute early, as in a normally distributed case would happen 16% of the time, the person would have to wait for the next bus to pass by. The opposite could happen if the expected waiting time was closer to the maximum, but since passengers are likely to arrive at the stop closer to the expected time, this is much less likely. [75]

Gathering the data for actualized timetables is relatively simple and can be done in several ways. As passengers cannot leave between stops, only the stops themselves need to be recorded. A stop detection method could be built based on vehicle speed. To filter out times when a vehicle stops at for example traffic lights, doors opening could be used instead, as that should only happen at places where passengers enter and leave the vehicle. However, some post-processing would need to be used as there might be fringe cases in door openings, such as something stuck in the door or simply opening the doors twice during a stop. A better but more manual method would require the driver to log the stops by something simple like a button press. However, this adds to the workload of the driver and is not entirely foolproof either. The best and most used method is location-based logging of stops. If the location of vehicles are constantly known, a dozen meters wide areas around the stops can be tracked, logging an event when they are entered and left. The two different timestamps are important, as they can show whether the driver stops for longer periods while he is early, bringing the actualized stop schedule closer to the planned one. This is of course not always an option, but the time spent on stops can also be used to measure how much is a late route due to stops versus driving. [15]

Safety, much like efficiency, can to an extent be derived from other sources than telemetry. Major incidents, such as crashes and collisions are logged outside the vehicle. The amount and type of these are visible from at least insurance claims, if not otherwise collected. While major incidents are arguably the most important metric, their sample size is, at least hopefully, too small to indicate minor improvements or decreases in overall safety. As such, more in-depth and minor safety metrics are required to measure safety more accurately. Minor incidents rarely leave a paper trail, since drivers are unlikely to report their own mistakes. A lot of minor incidents and close calls follow patterns that can be detected with telemetry. To detect these, however, general safety rules and standards are required to specify what kind of driving is considered unsafe. [13] [71]

Safety standards vary across the world, and they are very in-depth. For example, the European Union bus safety standards collection of regulations with just the references to the regulations themselves is 40 pages long [70]. Most of the regulations deal with the physical vehicle itself, but there are some driving instructions as well. Some obvious ones include the use of turn signals and stopping at pedestrian crossings. More intricate ones can set the thresholds for telemetry pattern recognition, such as maximum braking and not moving while doors are open. With a set of rules to build detection systems for, it is possible to measure safety not only based on the coverage of safety standards but also how well drivers follow them.

To detect deviations from safe driving, both simple thresholds of signals as well as complicated behavioral patterns can be used. For example, hard braking, which could be dangerous to people standing or moving within the vehicle, can simply be calculated from the change of speed. The derivative of speed is acceleration, but hard deceleration can still be safe if it is constant or slowly changing. Focusing on the derivative of acceleration, jerk, is more descriptive of the abruptness of braking. Similarly, sverving of the vehicle can be detected from the angle of the wheel with respect to speed. Outside of outright mistakes, aspects such as speed in turns, duration of turn indication and horn usage can paint a more general picture of the driving style. As the measurable aspects become more intricate, the safety standards no longer give a correct or optimal baseline. While this means that the driving style might not be comparable between cities, it is still comparable within it. The minor aspects of driving can be compared between iterations of routes as well as drivers themselves. If there are larger deviations from the average style, they can be pinpointed using the telemetry data and addressed to provide safer rides to passengers. [71]

Ecology is the overall sum of every part of the transportation system that strains the environment. These parts can have causation and larger correlations between them, causing the sum to be the most important metric rather than the individual parts. As such, every action taken must be judged by its overall impact, in case unforeseen consequences cause compensations in other areas. A tangible example of this would be the idling of a bus: keeping the engine running at a standstill needlessly consumes fuel, and buses spend a relatively large amount of time not moving. However, attempting to forcefully reduce the amount of idling and instructing drivers to shut down engines while at stops or even at traffic lights will not have the desired outcome. Starting the engine requires extra fuel and will also strain the engine more than idling would. If the engines last less, they need to be maintained or even replaced more often, costing more in both money and carbon emissions. The ecology aspect of driving does still have a lot of room for improvement in most cases, the negative feedback response just must always be considered as well. [72]

As mentioned in the previous chapter, the emissions and other strains on the environment should be divided by the provided passenger miles. From an ecological standpoint, the entire purpose of preferring public transportation is to reduce transportation emissions by sharing them between multiple people. While the total emissions need to be calculated, the accurate number of passenger miles is just as important a metric. Telemetry can easily measure the number of people that enter the vehicle. Whether by a cash register or a card reader, every payment for a ride is registered into memory. The method of payment and the ticket type are useful data in other areas, but the overall amount helps in calculating the ecology score. However, the passenger miles cannot be derived only from ticket purchases. Not all passengers are paying customers, as for example in Finland, children under 7, people in wheelchairs or a parent with a carriage do not need tickets in buses [76] [77]. Buying a ticket also does not necessarily reveal how long a part of the route is the passenger going to travel. More accurate measurements of passenger miles are discussed in Chapter 5.2.

The most immediate part of ecology is the consumption of fuel. Engines burn fuel to produce the energy to move, typically transforming gasoline into carbon dioxide and side products. The amount of carbon released is constant for a set amount of fuel, making carbon footprint calculations simple. For vehicles that consume 'clean' energy, such as electricity or hydrogen, the carbon footprint of manufacturing the fuel can be used instead. The manufacturing footprint should also be considered for fossil fuels as well, though. The amount of fuel a vehicle can easily be determined by the amount of fuel pumped into the vehicle during refills. For more time-sensitive measurements, it is also possible to constantly measure the amount of fuel in the tank, the change of which is the fuel consumption.

For longer periods of time, the deterioration of vehicles causes large carbon emissions as well. The lion's share of emissions happens during the manufacturing of vehicles, but maintenance and spare parts also add to it. As the vehicle deteriorates, it also produces more emissions during normal use [78]. For ecology calculations, emissions need to be averaged out for the entire lifespan of the vehicle. Much like the second-hand passenger cars discussed in Chapter 3.3, older vehicles are more prone to issues than new ones. The decision to replace the vehicles from the ecology's point of view should be when continuously repairing those issues becomes more costly in terms of carbon emissions. However, the breakpoints for comfort and more importantly money typically happen much sooner. There are no ready models to calculate vehicle wear for specific situations, but telemetry can help approximate it or find reasons for it. For example, if a vehicle mainly driven on gravel roads requires maintenance sooner and more often than an asphalt road counterpart, there is reason to believe that road dust wears the vehicle faster. This information could be used to circulate the roads vehicles use to even out the effect, resulting in the vehicle fleet staying closer in condition before they are replaced.

There are a huge amount of factors that affect fuel consumption and vehicle deterioration.

Many of these factors, such as the evenness of roads, cannot be addressed, but either planned or compensated for. Weather cannot be controlled, but the wear caused by humidity or coldness can be partly mitigated by for example storing the vehicles inside when not in use. There are also factors that can be addressed, such as driving style. In Chapter 2.2 it was mentioned that letting the vehicle 'roll' rather than brake while the clutch is up is a more ecological way of driving, as momentum is needlessly reduced. Many ecological driving recommendations also preserve the vehicle condition better, such as choosing the best gear to drive in. Other more individual methods of reducing emissions can be found via telemetry by comparing the fuel consumption of different iterations of routes. Exactly like in safety measurements, if the worst performances can be brought closer to the best ones by focusing on a few driving aspects, the overall performance increases. [11] [79]

Convenience is a subjective metric and cannot directly be measured by telemetry. There is no machine that could tell how much passengers are going to enjoy a ride. Telemetry can however measure the aspects of why or why not a passenger might have enjoyed the ride. Aspects such as temperature, noise and shaking are very easy and inexpensive to measure. For these metrics, there are no standardized optimal values, and the desired values are also not constant. For example, in colder seasons, people will wear warmer clothes and colder inside temperatures are needed to compensate for that. For sound, windy weather will bother passengers less than engine rustling would at similar decibel values. As context is required to best utilize this data, a person or a more complex machine learning algorithm is needed to provide it.

In the previous chapter, parts of convenience included features such as accessibility or Wi-Fi. Whether or not a vehicle has these features should be obvious based on if they are part of the model or separately installed. However, a vehicle can fully operate even without these features or if they malfunction. This means that a passenger boarding a vehicle cannot know for certain if the convenience features are available or not. Since the electronic systems to provide these features are quite complex, they either have or can be fitted with simple diagnostic properties. A telemetry system can with such diagnostic signals relay any malfunctions to appropriate places. The convenience features, the functioning of which can be considered binary, are a major part of overall convenience and are important for the owning company to provide either on-site or low-priority maintenance on as well as a random passenger that really needs to work online during travel.

Many outside services that provide convenience are connected to the vehicle telemetry. Electrical ticket purchases for example seats may leave out seats that are out of order. Applications and websites that track vehicle properties can show the previously mentioned availability of Wi-Fi, air conditioning and purchasable food items. Route finding and navigation software can use the live location of the vehicle to provide more accurate results. Some applications can even combine all of these together, such as taxi or Uber ordering applications. A vehicle or driver ranking is usually tracked within the software itself and not vehicles, even services using these remotely provide information about the vehicle. [42] [43]

## 3.6 Vehicle fleet telemetry

Of vehicles that operate repeating transportation routes, most belong to a business that owns multiple of them. The total of the vehicles a business owns, their vehicle fleet, can range from a dozen to hundreds of vehicles. Although, if a business does own hundreds of vehicles, they are often grouped to separate smaller fleets for clarity's sake. The management of multiple vehicles is more complex than managing a few vehicles would be, requiring either more time or more efficient methods. There are also many benefits to owning a larger fleet, such as a larger sampling of vehicle data and the ability to cycle tasks between the vehicles. [80]

In a small transportation company, the fleet manager might properly be able to keep track of less than a dozen vehicles without help. With close ties to the drivers themselves, any repairs can be reported in person and scheduled for a day, when fewer transportation orders need to be filled. Scaling up to a few dozen vehicles, some sort of management system is already required. The simplest and cheapest method would be a spreadsheet system, where each vehicle's work orders, schedules and drivers are listed for the employees to see. If a driver notices a fault in his or her vehicle, it can be listed in the spreadsheet for the manager to see. However, scaling up any further results in too much manual labor for one person to handle. Administrative tasks can to an extent be delegated to personnel responsible for a smaller set of vehicles each, but for the people responsible for the entirety of the fleet or the business itself, information needs to be significantly condensed.

Arguably the most important data to gather is the logistic aspect. The logistics contain where every vehicle is, what are they doing and who if any is driving them. If this information is known live, sudden changes can be reacted to fast and adapted to, whether another vehicle is needed to replace a broken down one, or a driver must be relieved due to a family emergency. The logistic information is not only useful at the current time, but historical data can be even more so. Knowing the total amount of orders completed in a frame of time can be helpful for analysis, or revealed past incidents can be studied with the logistic history, as it shows which vehicle and person were in question for a work order. The logistic data could be derived entirely without vehicle telemetry using work order plans as a base. However, telemetry can ascertain the truth, as it is able to log every part of data. The driver of the vehicle can be found out by software, for example by requiring a login before driving can be commenced, or by having a personalized electronic key that is used in the fleet's cars. The location of the vehicle can either be continuously transmitted using GPS data or only transmitted at arrivals and departures at key locations, such as drop-off points or stops. [80]

Beyond the tasks that generate revenue, namely the work orders, costs are also a very important metric to follow. The vehicle fleet itself will generate costs in three ways: fuel, repairs and storage. The fuel consumption, as mentioned in the previous chapter, can be continuously monitored in the tanks or logged at each refuel. The amount of fuel consumed by the vehicle relative to the work done in a frame of time can provide an approximate for the profitability of a business. Repairs are another cost that has to be accounted for, both in scheduled routine maintenance as well as actual repairs of broken parts. To gain a good understanding of the wear and future lifespan of the vehicle, it is typically checked by a mechanic regularly. It is both a more accurate way of finding out the condition of the vehicle as well as a safety-increasing method. Besides crashes and other sudden events that break down parts of the vehicle, most faults build up over time, and can at best be avoided by a tightening of a screw during scheduled maintenance. Having access to the history of repairs and maintenance done to the vehicle helps both schedule future maintenance as well as pinpoint aspects that might have led to extra repairs. Lastly, the storage of vehicles is not free. While it might be optimal for the vehicles to be operating constantly when not at maintenance, it is rarely a possibility. As such, the vehicles not in use have to be stored somewhere. If the business itself owns a storage hall big enough, the costs can be calculated by its upkeep costs, but especially when the business area is large, the vehicles have to be stored in a place that charges for it. The storage costs can be tracked by receipts received from these places, which is typically more accurate than trying to measure the storage time and location via GPS data. [81]

Besides the data required for day-by-day operations, a lot more can be collected with the intention of business improvement and optimization. For example, as mentioned in the previous chapter, the accuracy of the schedule is not absolutely vital for the work orders, but generally, it is desirable for everything to be on time. The people this kind of data is collected for are typically no longer within the management, but rather data analysts and scientists. The data analysts of course have use for the data mentioned above as well, but high variance data with underlying inobvious biases lead to wrong conclusions unless properly analyzed. The most useful knowledge extraction from data happens at both high and low levels. At a high level, the historical developments of practically every metric gathered are used to track trends and find reasons for any changes that might have happened. With low-level analysis, outliers and deviations in individual vehicles, routes or drivers can be pinpointed and worked on, providing focused and more efficient results.

For both high and low-level data analysis, the amount of data is key. With large enough sampling, the average and variance of every metric become more certain. A history of half a dozen vehicles over a week provides little besides the immediate revenue and costs. While it might be easy to find the vehicle, driver or route with the highest fuel consumption for example, the reason for it could be almost anything. Unusual amounts of passengers or cargo, heavy traffic due to roadwork or simply the driver having a really bad day that would not happen again are all in the realm of possibilities. Also, while business might have slowed down on Sunday, there is little reason to believe that business volume on the next day would differ from any other Monday, while the trend would suggest otherwise. However, if the sample size was much greater, for example a hundred vehicles over the span of two years, a lot more can be seen with certainty. Bigger events, such as new vehicles, routing or driver training can be seen in data, and their results can be evaluated.

# 4. TELEMETRY DATA APPLICATIONS

In this chapter, select telemetry applications are explored. These applications demonstrate both general and very specific uses of telemetry data to generate value. Businesses may have different approaches to these applications, so the business offerings themselves are referenced. However, as the exact implementations are trade secrets, the accessible information is limited. The home pages, blogs and other advertisements of businesses are taken with a hint of salt.

For clarity, telemetry data applications in this chapter are split into two groups based on their users. Companies can choose to enhance their business privately, holding the data to themselves. Alternatively, companies may share telemetry data in a controllable manner through applications that show its derivatives or the data itself to users. It is also not unheard of that companies would share some of their data for good publicity [82]. Either way, this is a good way to group the applications, as the approaches to generating business advantages have major differences depending on whether they are fully internal or not.

## 4.1 Private use of telemetry data

With a vast arsenal of telemetry tools, it’s left to the businesses to use them. Telemetry can be harnessed to optimize business and enhance services in an attempt to gain an edge over the competition or stay ahead of it. Business can be optimized by studying gathered telemetry data and making effective adjustments to improve different business areas. Service similarly can be enhanced by telemetry by providing easier use, informative choices or even clever advertisement. When one company makes an innovation, others are soon to follow, usually adding something of their own to it as well. Thus, the modern applications of telemetry are numerous and ever-increasing.

The default in business data handling is to keep everything possible secret. In the modern open data era this is often forgotten, but there is little business incentive in giving out data for free. First, maintaining open data sources costs money and time, especially if the data is updated regularly. The data must also be cleaned of any personal or sensitive information, or the company might face legal liabilities. Second, as mentioned in Chapter 3.6, more data is generally better for business analytics. If of two similar businesses only one makes their data open, the other business has twice as much data to analyze and benefit from.

Private use of data means that the data itself is never made visible to customers or the general public. This doesn't mean that the results of data use wouldn't be visible to the users, but is generally the opposite. With successful data use, rides can be improved to be more efficient, safer and with vehicles that are in better condition. For each respective attribute, the methods are business analytics, driver analytics and pre-emptive maintenance.

### 4.1.1 Business data analytics

The manners in which data analytics can produce value are numerous and dependant on the business. This is true in two ways: First, the amount and quality of data the business has collected correlate highly to the data analytics possibilities. Second, the business area and operations type limit the minimal magnitude of effective changes. What this means is that for example in transportation the true qualities of vehicle acquisitions are only apparent after some time. However, to obtain this data, a very expensive vehicle must first be acquired. Almost no example is shared between all transportation companies, so data analytics value must be considered on a broader scale. Generally, business analytics can be split into four areas [83]:

- Descriptive analytics (what has happened)
- Diagnostic analytics (why something happened)
- Predictive analytics (what is going to happen)
- Prescriptive analytics (what would happen)

Descriptive analytics uses data to describe past events. This is the least complex area of business data analytics and is actually mandated to a simple extent by law. Different countries have slightly different laws regarding this, but for listed companies in the EU for example it is mandatory to follow the international financial reporting standards, IFRS for short [84]. As the name suggests, these standards require reporting of the business's sources of revenue, profit and salaries to employees amongst other similar metrics. Even from these very basic metrics, key information can be found, from whether the values increase or decrease to whether the sources of income are probable or results of money laundering. [85]

For telemetry data, descriptive analytics provides mostly the same things it does for financial data. The measured metrics can be followed for their overall trends and the current situation can accurately be defined. The more interesting analytics follow if or when events or choices that affect metrics are studied. For example, if in an attempt to reduce fuel consumption or carbon emissions the fuel type was changed, whether the change was successful or not. This kind of descriptive analytics often works in conjunction with predictive or prescriptive analytics, as with them there is a clear hypothesis that can be proved or disproved in the future. [85]

Diagnostic analytics are used to clarify things, the exact reason for which is not known. In some ways the opposite of descriptive analytics, where the changes caused by an event are studied, diagnostic analytics can study what changes caused the event. If there is an unexplained rise or fall in a metric, other metrics are studied to see if any causation can be established. Since not all causation might be visible in data, it is typical to consult people with domain knowledge to either measure the root cause better in the future or look elsewhere for diagnostic needs. [86]

Diagnostic analytics can also be successfully applied to completely normal phenomena. For example, in general, it is completely ordinary for transportation to occasionally be late. Falling behind schedule simply happens sometimes and is completely out of the hands of the business. However, if each late route is closely observed, it might be possible to find reasons that if not caused at least contributed to the fact. Minor details that might cascade to minutes lost could include average speed, duration of stops or acceleration amounts. These metrics might be improvable, and focusing on them could allow transportation to follow the schedule closer. It is also possible that the lateness is caused entirely by outside factors, such as roadwork, a larger influx of passengers or an accident blocking traffic. This is also valuable data, as if the delays are caused by outside factors, maybe the timetable should be less strict for parts that are prone to delays.

Predictive analytics are used to predict the future based on the past. More accurately, what the future would look like in metrics if nothing out of the ordinary happens. More accurate approximations of the future come with many benefits, from expected finance flow to reduced warehousing costs. Predictive analytics also benefits from discoveries made in diagnostic analytics, as if a strong cause-and-effect is demonstrated, the effects the current causes will lead to can be expected. [87, p.22-29]

Predictive analytics can be done on a wide scale from daily approximations to yearly growth. The accuracy of predictions depends on the amount of historical data as well as its quality. Patterns from seasons to days of the week are more visible from a longer period of time and the expected variance in them can be calculated more accurately. Extrapolating from such repetitive patterns is fairly easy, but if larger changes have happened recently, it is more difficult to approximate their effect in the metrics. For example, if routes have been changed in either transport or delivery, it is harder to anticipate the recurring traffic impacts of holidays for them.

Prescriptive analytics is the final and most complex form of business analytics. Prescriptive analytics uses data to predict what would happen, if something that could happen, should happen. Unlike predictive analytics, which uses the premise that normal conditions would maintain, prescriptive analytics is based on hypothetical bigger changes that might happen in the future. Anticipating and knowing the boundaries for events that a business might have to react to allows preparation in advance. Historical data can be used to find similar events and study the effect they had on different metrics. [88]

Prescriptive analytics is not limited to changes that might or might not happen. Changes that will eventually happen also belong within the category of prescriptive analytics. A telemetry example of this would be for example the impact a newly built neighborhood has on the transportation network. It is apparent that routes passing nearby the new residents will see an increase in passengers. However, the magnitude of this is unclear beforehand. If the amount of passengers is too great for the existing transportation, the vehicles will end up full and people will start to prefer alternative methods of transportation. To more accurately predict the impact to the business, historical data can be used to approximate the number of passengers other similar areas currently house as well as their demographics.

### 4.1.2 Driver analytics

Be it a paintbrush, a gun or a vehicle, a tool is only as good as its user. While a person should be properly trained in order to use the latter two, skill discrepancies still exist even between trained people. The incentive to hire skilled individuals is apparent, as generally, people who are good at their job will provide more value to the business than those who are not. This value can be brought in a multitude of different ways ranging from work safety to more appealing business outlook. In a small office-based company the skill of workers is much easier to approximate, as more time is spent with them, and observing their work is easier. However, in a larger transportation company drivers perform mostly alone and have limited interactions with superiors or even colleagues. [89]

People are different and have individual strengths and weaknesses. In the context of driving, those with driving jobs typically have an aptitude for it, lowering the overall skill discrepancies. The driving skill level is not a single measurable attribute, however, but a sum of many different aspects of driving. It is unlikely that a person outperforms another person in every single category. Different driving jobs also benefit from different skill aspects by varying amounts. Whereas a long transport truck with an extremely heavy load is expected to navigate streets in a calm and slow manner, it is beneficial for a taxi driver to be more aggressive in order to move through city streets faster during rush hour. [90, p.118-129]

There are some aspects of driving considered to be more general and useful in most jobs and tasks. The most intuitive one is safety, as in no job is it beneficial to risk an accident. Of course, accidents can happen to anyone, and statistically such a rare event is more likely a coincidence than an indication of poor driving skills. As such, the detection of error-prone habits is a preliminary method to measure safety. Another general aspect of driving is the amount of fuel consumed. As discussed in Chapters 3.4.5 and 2.2, there are both driving habits as well as external factors that affect fuel consumption [11]. The upside of fuel consumption as a metric is that it is very easy to measure.

Companies approach maximizing the number of skilled drivers in different ways. The earliest method is in recruitment, where cherry-picking the best candidates is the common approach in almost every business. However, common metrics deemed beneficial in recruitment, such as job experience, might not reflect bad driving habits. Unfortunately, the aptitude and skill of a person for a job can only be accurately assessed after observation over a longer time period. Since people of different skill levels end up at a business regardless of recruitment efforts, skill maximization is focused on the two traditional methods: sticks and carrots.

A stick or the threat of one can be used to demand focus on key performance metrics. The core principle is simple: in an attempt to avoid the punishment, namely the stick, workers try to perform well. The figurative sharpest stick of course being the termination of employment. A slightly less aggressive threat is the docking of the next paycheck. However, these actions have a huge impact on workplace environment as well as the attractiveness of a business as a possible employer. Constant monitoring and punishments add pressure to workers and can even in some cases lead to worse performance due to extra stress. With the extremely high turnover rate in drivers, most workers would just seek employment opportunities somewhere else [91].

The extreme measures to lower bad performance are rarely used in modern businesses. While measures such as employment termination might have to be used to remove harmful employees, in countries with unionization and worker rights they are also strictly regulated. There are still some courses of action socially acceptable, such as reproaching by a superior or mandated extra training. These are generally viewed as unpleasant and still work under the same principle of people wanting to avoid them. Extra training in particular is a beneficial approach, as it works both as a deterrent for bad performance as well as ideally increasing the skill level of the trained driver. There are several businesses offering corporate partnerships for additional driver training. Some of the businesses even offer very specialized training, if exact problem areas are identified for the attendees. [92] [93]

A symbolic carrot can be offered by a company to incentivize better driving habits. The temptation of a reward can cause increased effort in measured areas if the reward feels reachable. The amount of possible rewards are low in the area of drivers, as the traditional reward for performance, career advancement, is limited. Also in the business perspective, those best at driving should stay driving, rather than for example managing drivers, which requires an entirely different set of skills. Since the drivers at its core are trading their time

Switching off your engine when parking for more than 5 minutes will increase your idling score.

***Figure 4.1.*** *A display of performance score and its five components [96]*

for money, the incentives are typically related closely to them.

There are multiple areas in which good performances can be rewarded. These areas can either be individually rewarded or they can add up to an overall score. A US-based semi-truck delivery company MTS, Martin Transportation Systems, for example, has an individual reward for perfect roadside scaling results and an overall reward for driver performance. Highlighting key areas of significance or improvement with individual rewards will add extra focus to them, but an overall score has less variance in it caused by occasional mistakes that might happen to anyone regardless of skill. An overall score can be composed of as little as two metrics, as is the case in another semi-truck company Decker [94]. Much like in Figure 4.1 or MTS reward program, a wider array of metrics is typically considered in the overall score. [95]

Balancing the prizes and rules of a reward system is imperative. There are two ways to set a rewarding system threshold: a common point goal or a relative threshold. If the common point score, either of an individual metric or of a combined total, is beat, a reward is granted. This is the more common system but does require extra effort in setting the goal. The easier system to implement is a relative threshold, where the top percentile of scores are granted the reward. This does not require extensive metric analysis and also the cost of the reward system to the business will be constant. The monetary amount of rewards has to be balanced so that the rewards feel significant to the drivers but aren't too costly to the company. In Table 4.1, several performance bonus systems and their maxi-

| Company | Score type | Fuel (MPG) | Safety | Operational | Pay increase |
|---|---|---|---|---|---|
| Martin [95] | Combined total | Yes | Incidents | Yes | ∼ 4% |
| Carter [98] | Combined total | No | Behavioural | Yes | Flat 3% |
| Schneider [99] | Individual goals | Yes | No | Yes | ∼ 9% |
| Melton [100] | Combined total | Yes | Both | Yes | ∼ 9% |
| Nussbaum [101] | Combined total | Yes | Both | Yes | ∼ 10% |
| Decker [94] | Combined total | Yes | No | Yes | ∼ 8% |
| Bison [102] | Unknown | Unknown | Yes | Unknown | ∼ 19% |

***Table 4.1.*** *A comparison between 7 US-based trucking companies' driver bonus systems*

mum rewards for US-based trucking companies are listed. As driver pay is not constant between companies, unknown pay is defaulted to an average driver earning 0.50 dollars per driven mile, driving 2500 miles per week [97]. [96]

From Table 4.1 one can see trucking companies place the maximum amount of bonuses below 10% of overall pay. Intuitively, this seems to fall between an insignificant and too significant amount of money. The majority of companies employ a point total for bonuses, allowing drivers to compensate for their weaker areas with their strengths. In addition to fuel consumption as an important metric, there is a significant focus on operational metrics. Tardiness, completed work orders and failed weightings are common metrics that do not require vehicle telemetry. Similarly, while there are some behavioral scoring systems for safety, they are less represented than penalized incidents. [102] appears to be an outlier, as very little information is provided about the calculation for bonuses. The considerably high maximum bonuses might be entirely unreachable for normal driving. The bonuses also incentivize hyper-optimization of metrics and may even lead to blatant cheating by clever drivers.

The competitive integrity of reward systems needs to be upheld. As discussed in Chapter 3.4.5, purely optimizing single metrics can end up causing more damage in other, unmeasured areas. Attempts to cheat might be able to be spotted from data if for example engine is shut off often to save fuel or undocumented refills are used to cheat the measured amount of fuel consumed. However, tighter observation of drivers is not kindly looked upon, so the problem is better tackled in the reward system itself. If the reward is easy enough to reach by simply driving well, it will be easier to just perform better than to cheat. Therefore high competition by relative thresholds and large rewards for near-perfect point totals are uncommon in the industry.

### 4.1.3 Predictive maintenance

Every machine will in the course of time eventually break down. When a machine breaks down, there are two ways to continue operations: acquiring a new machine or repairing the old one. Both of these options however cause downtime in operations, and eventual breaking down is difficult for a business to accurately plan for. This is why before the eventual breaking down, most vehicles undergo planned maintenance, where smaller repairs can be made during scheduled downtime to maintain operations. The condition of the vehicle can also be observed at each maintenance in order to plan for future maintenance needs. [103, p.2-3]

A vehicle can break down in a very high number of different ways. The man credited for the famous Murphy's law, Edward J. Murphy, was an airplane engineer by occupation [104]. While the ground vehicles this thesis focuses on have less critical systems than aircraft, every one of those systems will eventually go wrong, as they can go wrong. It is important to identify common critical failures and their causes in order to get the maximum effect from periodic maintenance. While some failures in parts can be abrupt and unpredictable, typically failures become more common the more parts wear down.

The time it takes for part failures to happen is not constant. A part with the expected lifespan of half a year in normal use might last anywhere from a month to years. From parts that have previously failed, a time distribution for part failure can be approximated. If no considerable amount of parts fail in the first three months, the part in question does not require a scheduled maintenance before that. But if for example 5% of parts fail before the fourth month, the failures are common enough that the parts should be inspected by then. The interval between scheduled maintenance is an optimization function between the cost of maintenance, the cost of failure and the approximated amount of failures. [105]

Since the amount of wear is not constant, some vehicles might go under maintenance without need and part failures could've needed maintenance sooner. There is a big monetary incentive to predict the need for maintenance. Predictive maintenance can both reduce the number of unneeded maintenance as well as the amount of part failures. From data that the vehicle gathers, the condition of parts can be approximated without the physical inspection. There are two ways to predict the need for maintenance: observing the data for symptoms of wear and generating a wear profile from use history. [105] [106]

Symptoms of wear before failure can be extracted from data in many ways. Some of the possible methods were introduced in Chapter 2.5, namely correlations, patterns and anomalies. A correlating symptom would for example be a leaking tire. As air escapes the tire, its pressure will fall, causing increased friction and skewing of the car. These symptoms correlate most to the pressure of the tire, but if that is not recorded, it might also be seen from the average wheel position. Spotting a leaking tire before the tire is completely flat can save a car from an extra stop in a bad location.

Patterns and anomalies are more complicated ways to discover wear. Correlations are easy ways to spot if something is wrong, but it can be hard to pinpoint the problem and assess its severity. Increased fuel consumption for example could indicate a number of different errors from engine problems to deflated tires. If the measured metrics correlate during a set of specific actions, a pattern can be established instead. For example a specific problem might be visible during low speeds and RPM in acceleration, but not anywhere else. Anomalies in telemetry data typically suggest bigger problems but are prone to measurement failures. If the vehicle has transmission issues shifting gears, abnormal changes in speed would be a symptom of that. However, if a speedometer fault causes readings to randomly spike, it might cause very similar results. [103, p.13-19]

Predicting vehicle wear from historical data is not an exact science. As discussed in Chapter 3.3, approximating the condition from metrics such as odometer reading may give some idea about the condition, but not the whole picture. However, with a large vehicle fleet of similar vehicles with different amounts of wear, data can be used to deduce indicators that one vehicle would suffer more wear than another. The research paper in [105] also wasn't limited to signals the car provides such as mileage, but analyzed more indicative data from previous maintenance visits. The results were not surprising, as cars that had previously been in worse condition were more likely to be so again. [106]

## 4.2 Public use of telemetry data

Large-scale telemetry data is not limited to just private businesses. There are areas of business interest where it is profitable to give data away. Some services, like transportation live location, enhance the user experience of using respective services. Some companies use their customers for data, and share everyone's anonymized data with all customers. Customers can then compare themselves to other drivers based on vehicle model or geographical location. [107]

Location is an important telemetry data point that is useful on many magnitudes. Even at very low accuracy location can be used to locate the country the vehicle is in. This can allow for example the first-time setup language and traffic rule detection. At slightly more accurately, correct weather information can be displayed in the vehicle, or a package can be tracked at city level by the recipient. At pinpoint accuracy, a vehicle can be found in a parking lot. The usefulness of location is also highly affected by its latency. A package traveling city by city can be tracked by the hour, but if the package is delivered to the recipient's doorstep, the vehicle's location is desired live.

With modern technology, it is extremely simple to broadcast one's location. For exam-

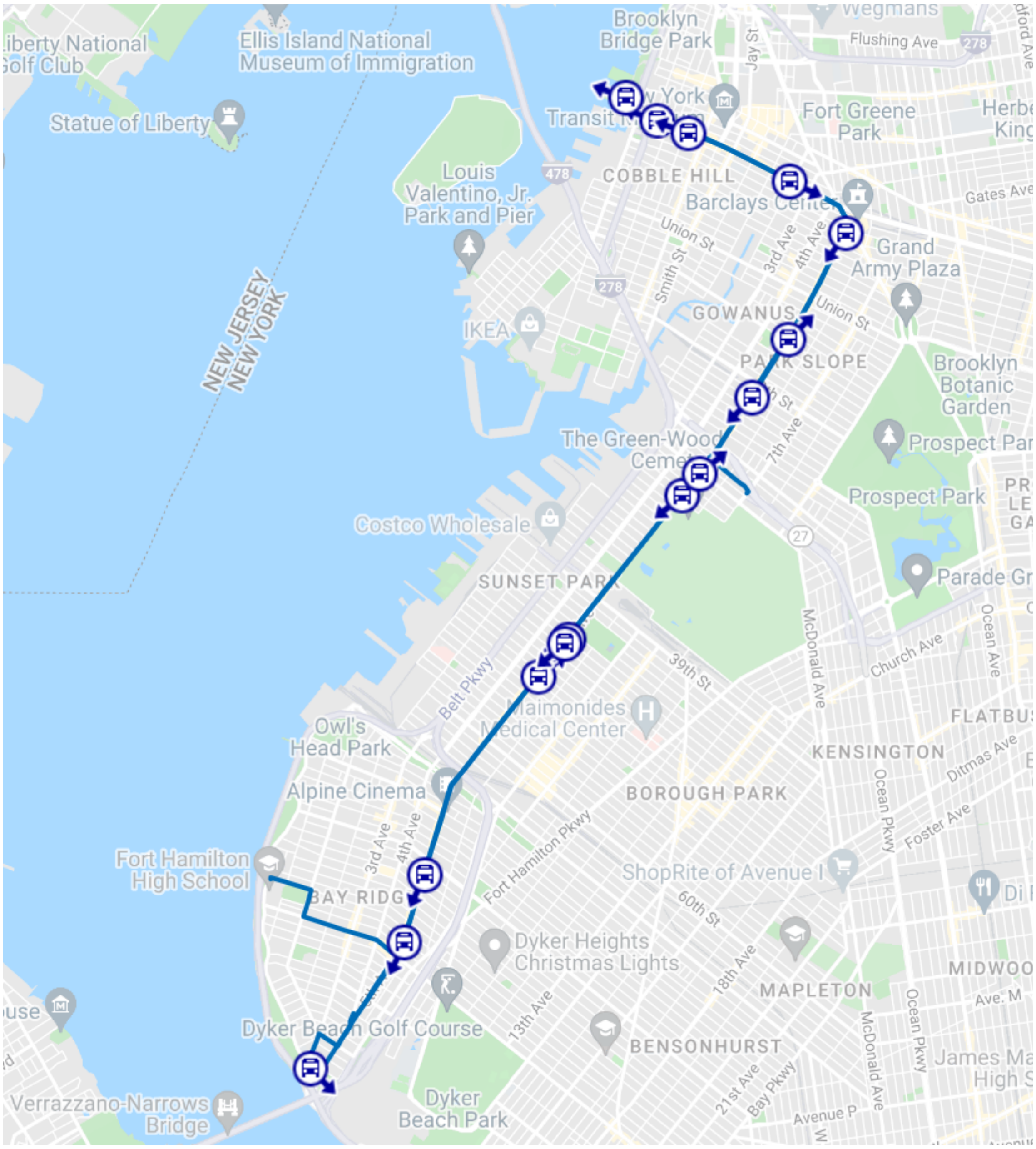

***Figure 4.2.*** *Live view of buses on route B63 in Brooklyn, NY [15]*

ple, each smartphone is capable of this. Many smaller company employees use one of several dozen possible smartphone GPS applications to obtain or broadcast location information. These applications are most commonly used for navigation purposes but have the possibility to broadcast location to the internet if live tracking of deliveries for example is required. Larger vehicles and modern cars can have GPS-capabilities built-in instead. Through the internet, it is then possible to observe the vehicles on a map, as shown in Figure 4.2. [74]

There are several modern services that use live location. Almost all transportation businesses have a live tracking map: public transportation can be followed to better schedule traveling, waiting taxis can be found from a map, also packages and deliveries can be tracked in order to be ready to receive them. An individual may wish to track their vehicle in order to find it in case its location is forgotten or it has been stolen. Another common use is navigation software, which can provide correct and updating driving instructions through live location. [74]

Live location can also be linked to other metrics. When choosing a method of transportation, there can be more deciding factors than just availability. Of course, rural areas are typically limited to the closest bus routes regardless of the condition of the bus, but urban areas might have multiple options for traversal. If telemetry data, such as inside temperature, is linked to the bus tracked on a smartphone app, it can make a large impact on the decision-making. During the summer it is likely for a person to rather choose a vehicle with working air conditioning for the cost of waiting extra time. Another metric affecting travel comfort is fullness, as a packed vehicle is less enjoyable than one with empty seats. Vehicle fullness is discussed further in Chapter 5.2.

The vehicle telemetry applications described for private use can to some extent be used by people as well. The number one problem in telemetry applications for an individual is the limited sample size available. With just a vehicle or at best a couple within family, only inaccurate, relative conclusions can be drawn. There are some practical but niche use cases, such as accurately splitting the upkeep costs of a shared car between its drivers. However, the only way to get an appropriate sample size is to share data to a service, which can combine the data of individuals into one big sampling.

The companies most likely and willing to obtain their user's data are car manufacturers. First of all, it is very easy to get customers to agree to the exchange of telemetry data amongst other things when the vehicle is purchased. Second, the data customers provide are very uniform, as all vehicles have the same manufacturer, the said company. Lastly, the car manufacturer can largely benefit from customer data besides sharing it back. While the manufacturer builds their own driver ranking application, they can receive habitual tendencies and difficulties the users face driving their vehicles. This, otherwise unavailable data can then be accounted for in the design of new models. Similarly, if there is a diagnostic predictive maintenance application, the manufacturer can receive data about the fastest parts to wear down and attempt to improve them in future models. [106]

# 5. FUTURE OF VEHICLE TELEMETRY

Vehicles and the telemetry data they gather have advanced considerably during this century. The exponential growth of processing power and increased connectivity allows the functioning of the modern applications introduced in the previous chapter. With the new rise of machine learning, the analysis and use of motor vehicle telemetry data will continue to develop. This development will introduce entirely new innovations as well as deepen the current applications of telemetry data.

This thesis will focus on one example in each development area. With the first prototypes already cleared for road use, autonomous cars will undoubtedly increase in number during the following years. These vehicles require a lot of additional measuring equipment in order for the artificial intelligence to drive, and thus provide additional data for analysis. Additional measuring equipment can also be added to vehicles for the sake of data analysis only. Public transportation has an incentive to better their understanding of the efficiency of each iteration of a route. The number of people inside can be better tracked by adding new sensors or adjusting current ones.

## 5.1 Autonomous vehicle telemetry

Autonomous vehicles are vehicles that can navigate without an operator. Operator-assisting has existed for a long time and is most commonly known in airplane autopilots and cruise control. If not strict with the definition of a vehicle, warehouse robots on tracks have achieved the status of an autonomous vehicles decades ago. However, operator-assisting vehicles still require an operator, and warehouse robots exist in a well-controlled environment. Vehicles that can transport people autonomously have to be extremely safe, and the technology is not entirely there yet. [108]

The easiest vehicles to turn autonomous travel on tracks. Trains, trams and subways do not need to worry about steering, as their heading is constant. However, there are only a few of these vehicles without an operator in the world. While an operator is still required, the role is mostly supervisory. The operator ensures that the speed of the vehicle remains within safety limits and closes the automatic doors when people are done entering and exiting. These are both relatively easy tasks to automate. The reason these vehicles are not autonomous yet is the relative low cost of operators to the development of autonomous systems of the same degree in safety. [109, p.26-29]

Steering is not the only factor that makes autonomous cars more difficult than trains. Roads as an environment are much more chaotic than tracks, as there are other vehicles and pedestrians in traffic. The other vehicles also might not be autonomous and therefore the driving AI cannot communicate with them, but must instead react to their unpredictable actions. Pedestrians can also behave in an unpredictable manner, and in rural areas, wildlife is guaranteed to act so. Reaction speed and the ability to detect potential dangers are the two most critical problems in autonomous cars. [110]

A regular car cannot be taught to safely operate autonomously. Older car models have no way to identify their location or stay on road. Newer car models, with highway cruise control and GPS, might be able to follow relatively straight roads at low speeds safely, but they are unable to react to uncontrolled situations or make tight turns. There are several key components and upgrades that autonomous cars require that differentiate them from traditional cars. Autonomous cars require an accurate model of their surroundings, which can be built with sensors including but not limited to radar, visual and sonar. To process the data into the model and make decisions in traffic quickly, significantly more processing power is required than what would be available in any traditional car. [108] [111]

With additional sensors, the possibilities of telemetry use beyond navigation increase. While safely driving remains an autonomous vehicle's number one task, additional gathering of data should not be neglected. The additional sensors provide additional data sources for analysis. The new sources, such as the mentioned radar, visual and sonar, are alone not that useful for traditional insight. However, they are valuable when driving AI development is considered. The "thinking process" of video processing can be stored from situations where a wrong conclusion was reached to better tune the algorithm in the future. This is what happened in 2018 when the first pedestrian death by autonomous vehicle occurred [112]. The vehicle stored as much data as possible from the time around the accident, which provided the conditions for wrong conclusions and the reason for the accident. If autonomous vehicles are able to learn from their own mistakes as well as other vehicles', the model will grow more robust. [113]

Additional sensors are not limited to autonomous operation, but internal telemetry as well. The ability to notice and compensate for changes in the car's handling is easy for a human driver, but not for an AI. For example, if the effectiveness of brakes decreases, a person will simply press the pedal harder when he subconsciously notices the car isn't slowing down fast enough. For an autonomous system, the braking amount has to be adjusted beforehand to ensure smoothness in slowing down. This example generalizes to other minor faults caused by wear as well. The driving feel discussed in Chapter 3.3 as a means to approximate a vehicle's condition is unavailable in autonomous vehicles. Excluding faults noticeable to passengers, such as weird noises, the need for maintenance has to be predicted by the vehicle itself. [114]

## 5.2 Public transportation fullness measuring

The future of telemetry does not only include new innovations but a deepened measurement and understanding of current metrics as well. One of these metrics is the ecology of public transportation. As explained in Chapters 3.5.3 and 3.5.4, ecology is best measured in carbon emissions over passenger miles. However, while the carbon emissions are quite simple to approximate, passenger miles are more difficult. If the amount of passengers in a vehicle, its fullness, is known at all times, ecology can be measured accurately. With the ecology metric more accurate, routes can be compared to each other and new routes can be designed more effectively. [68]

The benefits of accurate passenger miles calculation differ in value depending on the method of transportation. While more accuracy in data is never detrimental, the improvement in data is not actionable for track-operated vehicles. Building new tracks, especially underground, is extremely costly, so the decisions that lead to new tracks are not easily swayed by improved ecology estimates. However, some of the possible solutions to fullness measuring benefit trains, trams and subways as well. Buses are the main beneficiary of fullness measuring, as their routes can be altered to go everywhere roads do.

Buses have sensors unique to them that can assist in fullness measuring. A common feature, even in older bus models, is the ability to lower or "kneel" the vehicle at stops by slightly deflating stabilizing airbags above the wheels. Kneeling is especially helpful for the elderly, people in wheelchairs and people with strollers, as it makes boarding the bus easier. After entering and exiting the bus have ended, the bus inflates the airbags back to normal. The force required to undo the kneel is relative to the amount of weight the passengers cause. The amount of force can be measured to approximate the weight change during each stop. However, the weight of passengers is relatively small compared to the weight of the bus, and a larger adult might weigh as much as five children. The method is thus heuristic by nature, but it is cheap to produce and does not require additional sensors to be installed. [115, p.10-14]

Another method of counting the number of passengers inside the vehicle would be to log the amount of entering and exiting. A pair of sensors, typically optical, can determine if something is passing between them. Since due to the operating environment, something can always be expected to be a person. If at least two pairs of sensors are installed to each of the doors, the direction a person is heading can be determined by the difference in their activation. This system is however prone to mistakes if multiple people can fit between them at once, as it is impossible to determine how many people are between the sensors. [116]

Modern buses, trams and subways have a live camera feed of doors. Since it is a driver's responsibility to close the doors once everyone is done passing through them, cameras allow them to do so safely. From this live video, it is possible to use edge computing and machine learning algorithms to detect people and their movements. This technology is useful for track-operated vehicles However, much like for the optical sensors, occlusion is also an issue for cameras. A less error-prone method would be to detect the number of passengers in the vehicle instead of the net change. This does require extra cameras to cover the entirety of the vehicle. With AI, it is then also possible to track every passenger from entering to exiting. This allows refining the data further to show the average number of people that travel from one stop to another. [116] [117]

There are privacy concerns linked to camera use and tracking. People generally value their privacy and do not want their traveling to be tracked. While details about the individual can be left out and only anonymized data kept, the possibility of wrongful use remains the main deterrent. However, there is a possibility to use another type of optical camera to ascertain the anonymity of passengers. People cannot be identified from thermal camera footage, but their movement and number can still be tracked. Unfortunately, thermal cameras are more expensive and bear the outlook of normal cameras, which leaves people thinking they are being recorded on regular video. Thermal cameras also are prone to mistakes, if the surrounding area is especially hot, which is something a bus during summer can be. [118]

The usefulness of fullness measuring is not limited to ecology calculation. The historical data for analysis is mostly limited to it, but there are also uses for the live data. Much like live data for location used in transportation, the passengers can benefit from this data while planning their trips. If the data is made public through the live location apps, a person can choose a bus that is not almost filled to legal capacity. This benefit is limited to urban areas where there are multiple route options. Fullness as a metric is also prone to quick changes, as a class of schoolkids might enter the bus just one stop before a person would, leaving very little room for reaction.

# 6. DISCUSSION

In this chapter, the main topics and findings of this thesis are discussed. A more practical approach is taken in the evaluation of topics covered in previous chapters. Key themes, problems and solutions are raised and their surrounding contexts analyzed. The amount of unavailable information is considered when making conclusions, but presented solutions are more general in nature.

## 6.1 Measured signals

The oldest measured signals are still the most prevalent. Operation critical signals have been signaled to the vehicle operator for decades. More signals have been added in the course of time, but almost none have been removed. The use of some signals has evolved with automatic assisting systems, such as ABS and automated parking. Instead of showing the fact that some tires are sliding, the ABS activates and ABS activation is displayed instead. However, as the system can be disabled, sliding signals can still be required.

There are several problems regarding the addition of new measured telemetry signals in vehicles. The first problem is the aforementioned fact about few signals being removed during the passage of time. Vehicle development is incremental, where in order for new models to sell well, they need to have all or at least most of the previous functionality as well as improvements. As was seen in Figure 3.1, there are already several dozen signal lights, far more than an average driver could keep track of [48]. Any additional signals are sure to be lost in the group and see less use than they would if the dashboard was more streamlined.

The second problem with new signals is that they need to provide short-term, tangible value. The circular analytics process as explained in [31] is severely hindered if the circle can only be started from a functional application. While some vehicle manufacturers have test models in laboratories, larger sample size can only be accumulated from sold models. Not only can gathering data from sold vehicles be difficult, but it is also difficult to sell measurement signals that only benefit the manufacturer.

The final problem is more general and is caused by the high cost and long lifespan of vehicles. The cost of new vehicles is measured in tens of thousands of euros. These vehicles are expected to be functional for over a decade, or possibly less if they are later sold forward. It would cause a business significant financial stress to acquire an entirely new fleet of brand-new vehicles. In reality, as the business requires more capacity or older models break down, newer models are bought to fulfill the need. As such, the vehicles are relatively old on average, and models that can be considered new form only a fraction of the fleet. Any models that contain measurement for signals that could be used for business analysis and such will require as long as years to cover a large enough part of the fleet.

There is a partial solution to all three problems: external measuring equipment. External measuring devices only require power from the car's battery or they can even be battery-powered. Fully external measuring is also possible in cases such as traffic counting and weigh-in-motion [62]. Depending on the device, it might require some open connection port from the vehicle, but more general devices can be installed on even decades-old vehicles. These telematics devices can range from a GPS to dash-cams to inter-engine signal storage. A more modern method is to have the driver's smartphone act as a telematics device. Applications such as [73] do not need to require car models with built-in GPS capabilities and worry about its communication to their systems when a simple programmable interface in a smartphone can suffice. [119]

## 6.2 Telemetry applications

When discussing telemetry applications in Chapter 4, a split between private and public use of data was made. Furthermore, the topics were confined to general business analytics and two actual applications. There are two misconceptions caused by these limitations: First, the telemetry data applications would be limited to a select few, and second, the difference between public and private use is clear.

The majority of telemetry applications are area-specific. Even the more general applications such as the discussed driver analytics and predictive maintenance are rarely implemented as-is to an area. If the vehicle is used for anything else than driving, that unique activity will also have to be tracked. Specialty trucks such as firetrucks and garbage trucks are good examples of this, as measuring water pressure in a fire truck is very useful, but other areas have no use for it. For garbage trucks, some entirely different applications are needed instead to track the loading and unloading of the truck [120].

The differences in implementation are not only due to slight differences in business areas. In private business telemetry applications, the businesses have to own and therefore develop their own versions of applications. Even if the telemetry solution is bought from a third party, there is competition between providers and products. Much like the slight differences in approach visible in Table 4.1, companies are likely to have unique solutions even if they are publicly advertised. As such, the differences can be expected to be even more different in solutions, which aren't publicly available for this thesis to compare.

Telemetry applications are rarely entirely private. For one, if any data collected for an application can be linked back to a person, by data protection laws he must have possible access to it [121]. Also, people are generally unlikely to give their anonymized data away for free. For an application to be entirely private, the sample size needs to be limited to people in the company. As most companies are not large enough to gather enough data by themselves, they need to share their data with either other companies or the general public.

Innovations in data trickle down. Contrary to many other areas of business innovations, larger businesses spearhead the business. Since data is required to develop new applications, it is only possible for those that have access to it and are willing to analyze it. The first draft of many modern applications were proofs-of-concept by larger businesses, after which smaller parties refined their own, more public versions. For example, the discussed driver analytics started in private big trucking businesses and has recently appeared to the public through car insurance businesses [122].

Driver analytics is an evolving telemetry application, as is predictive maintenance. In addition to the solutions being different, none of them are perfect either. In driver analytics, the simplest metrics are very prevalent in solutions. While less fuel consumption and idling are generally good, if they are the only optimized metrics, unhealthy operating habits can form to "game the system". Similarly, the metrics in predictive maintenance can lack direct causation. Previous repairs, their costs and vehicle age were used in [105] and provided good results, but the approach is still very heuristic in nature as none of those metrics cause breaking down by themselves.

## 6.3 Vehicle telemetry's future

To predict the future with predictive analytics, past data is explored for trends and previous events. Following this approach, the future of vehicle telemetry can also be predicted. The development and new signals in vehicles stagnated for a longer period of time before the recent AI boom. An AI winter has always followed a boom, but is currently nowhere near. The surrounding hype has slightly diminished though, partly due to recent world events. Therefore the continued trends are expected to continue for at least a few more years and depending on their success, possibly longer. [123]

The current trends in vehicle telemetry are more extensive data gathering and vehicle interfaces to linked devices. Business intelligence has become more dependent on data so the improvement in both data volume and quality is ongoing. The newest car models have accompanying smartphone applications and IoT-properties [124]. The additional connectivity also aids in data collection, as wireless transmissions of measured signals are made easier.

For future applications, new innovation and a possible improvement in autonomous vehicles and fullness measuring respectively were explained. These two applications have in common a strong link to machine vision. Accurate enough object detection and recognition software has become present in the last few years and is no longer preventing these applications. Algorithms such as [117] and especially their recent versions [125] are able to run on light edge devices and are free to use.

When asked about the biggest challenges in vehicle autonomy, experts agree on safety as the largest concern. From a technical standpoint, [113] and [126] list weather conditions and radar interference as the most significant current shortcomings. Bad weather does not only reduce visibility but inhibits other signals as well. Radar interference is critically detrimental in two ways: First, radar is one of the primary tools for detecting other vehicles. Second, for partly autonomous vehicles that are remote-controlled, it severs the connection and the vehicle is stranded. If autonomous vehicles are susceptible to accidents, which they to an extent always are, the focus shifts to the party legally responsible in that situation.

The largest limiting factor in vehicle autonomy is legal liability. [110], [111], [113], [126] and [127] all agree that before autonomous vehicles become mainstream, the legal stature for autonomous vehicles must be established. Eventually, insurance companies will form partnerships with fully autonomous vehicle manufacturers, solving this issue. However, in the meanwhile, the legal position and responsibility of a supervising operator are questionable. [126] claims that since operators are not in constant control if the autonomous system drives the operator to a position where there is no time to react to or avert the accident after control is transferred, the operator cannot be considered responsible.

Legal liability also causes skepticism in the general public, discouraging potential buyers. Businesses are rightfully hesitant to sell autonomous vehicles when a minor fault could lead to massive lawsuits. Similarly, potential customers do not want to purchase an autonomous vehicle, the behavior of which no one wants to take responsibility for [127]. [111] even ranks social acceptability as a bigger challenge than safety and legality. The research and development process of autonomous vehicles is extremely costly to vehicle manufacturers if they cannot sell any intermediate versions.

# 7. CONCLUSION

This thesis studied the theory and applications of vehicle telemetry. Both immediate signals and long-time history analysis were considered. Private and public use of telemetry data were introduced for singular and fleets of vehicles. A comprehensive look at the past and modern capabilities, as well as two future applications, were explored.

No vehicle can safely operate without telemetry. This was established with the introduction of immediate telemetry signals, as data such as current speed is required to drive a vehicle. The scope of immediate signals was later broadened to include data a vehicle can transmit outside. The data is most commonly about the vehicle's location, but can also include inside temperature, or even its fullness.

Historical data can be built from recorded immediate telemetry signals and other metrics. Uses for this data were shown in the form of applications and performance metrics. Performance metrics for public transportation, efficiency, safety, ecology and convenience were explained and analysis use cases were exemplified. Current applications were discussed in private and public use cases.

Two modern private data use cases, Driver analytics and predictive maintenance, were explored with examples. Businesses that either use or offer these services were used as references. For driver analytics, different types of driver rewards systems and metrics by which they are granted were discussed. With predictive maintenance, key focus areas were considered. Methods to detect wear were introduced and possible future improvements were discussed as vehicles become more autonomous.

A future innovation and analysis improvement were discussed in the form of autonomous vehicles and vehicle fullness measurement. Autonomous vehicles, while somewhat present already, are rare even in cases where autonomy is relatively simple. The safety reasons and other challenges leading to this were discussed as well as improvements required for modern vehicles to perform autonomously. New sensor types and possible telemetry analysis use outside autonomous navigation were introduced. Vehicle fullness was explained as a problem, and several possible solutions were discussed. Lastly, possible benefits of making the data available to the public were discussed.

# REFERENCES

[1] Ertug, I. *Odometer manipulation in motor vehicles in the EU*. `https://www.europarl.europa.eu/RegData/etudes/STUD/2018/615637/EPRS_STU%282018%29615637_EN.pdf`. Cited 23/05/2022.

[2] Merrian-Webster. *Telemetry*. `https://www.merriam-webster.com/dictionary/telemetry`. Cited 13/07/2021.

[3] Merrian-Webster. *Telemetering*. `https://www.merriam-webster.com/dictionary/telemetering`. Cited 13/07/2021.

[4] Cambridge dictionary. *Telemetry*. `https://dictionary.cambridge.org/dictionary/english/telemetry`. Cited 13/07/2021.

[5] Princeton Wordnet. *Telemetry*. `http://wordnet-rdf.princeton.edu/lemma/telemetry`. Cited 14/07/2021.

[6] Wiktionary. *Telemetry*. `https://en.wiktionary.org/wiki/telemetry`. Cited 14/07/2021.

[7] Miller, P. *What is edge computing?* `https://www.theverge.com/circuitbreaker/2018/5/7/17327584/edge-computing-cloud-google-microsoft-apple-amazon`. Published 07/05/2018.

[8] Australian Transport Safety Bureau. *Black box flight recorders*. `https://www.skybrary.aero/bookshelf/books/3679.pdf`. Cited 14/07/2021.

[9] Corporate Finance Institute. *Moore's law*. `https://corporatefinanceinstitute.com/resources/knowledge/other/moores-law/`. Cited 15/07/2021.

[10] Backblaze. *The Cost of Hard Drives Over Time*. `https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/`. Cited 15/07/2021.

[11] Ecodrive. *The golden rules of ecodriving*. `http://www.ecodrive.org/en/what_is_ecodriving/the_golden_rules_of_ecodriving/`. Cited 18/07/2021.

[12] Sumo Logic. *What is a log file?* `https://www.sumologic.com/glossary/log-file/`. Cited 18/07/2021.

[13] DriveRisk. *4 essential safety metrics you should be tracking*. `https://driverisk.com.au/4-essential-safety-metrics/`. Cited 06/06/2022.

[14] Harrisonburg Traffic Ticket Law Firm. *What is a Velocitation?* `https://www.keeferlaw.org/what-is-velocitation/`. Cited 20/07/2021.

[15] MTA Bus Time. *Live location of buses on route B63*. `https://bustime.mta.info/#B63`. Cited 11/05/2022.

[16] Harel, K. *Self-driving cars must be able to communicate with each other*. `https://ingenioer.au.dk/en/current/news/view/artikel/self-driving-cars-must-be-able-to-communicate-with-each-other/`. Cited 24/07/2021.

[17] Talend. *What is Data Redundancy*. `https://www.talend.com/resources/what-is-data-redundancy/`. Cited 24/07/2021.

[18] Amazon. *Amazon Web Services*. `https://aws.amazon.com/`. Cited 25/07/2021.

[19] Microsoft. *Azure*. `https://azure.microsoft.com/en-us/`. Cited 25/07/2021.

[20] Google. *Google Cloud*. `https://cloud.google.com/`. Cited 25/07/2021.

[21] IBM. *Benefits of Cloud Computing*. `https://www.ibm.com/cloud/learn/benefits-of-cloud-computing`. Cited 28/07/2021.

[22] WINZIP. *File Compression: Everything You Need to Know About Compressing Files & Why*. `https://www.winzip.com/en/learn/tutorials/file-compression/`. Cited 28/07/2021.

[23] AWS. *DynamoDB*. `https://aws.amazon.com/dynamodb/`. Cited 28/07/2021.

[24] AWS. *EC2*. `https://aws.amazon.com/ec2/`. Cited 28/07/2021.

[25] AWS. *S3*. `https://aws.amazon.com/s3/`. Cited 28/07/2021.

[26] AWS. *Lambda*. `https://aws.amazon.com/lambda/`. Cited 28/07/2021.

[27] AWS. *Glue*. `https://aws.amazon.com/glue/`. Cited 28/07/2021.

[28] AWS. *Sagemaker*. `https://aws.amazon.com/sagemaker/`. Cited 28/07/2021.

[29] Berman, R. *Infographic: How dangerous is your daily commute?* `https://bigthink.com/robby-berman/reasons-to-take-the-train-to-work-instead-of-dying`. Cited 29/07/2021.

[30] National Safety Council. *Drivers are Falling Asleep Behind the Wheel*. `https://www.nsc.org/road-safety/safety-topics/fatigued-driving`. Cited 29/07/2021.

[31] Schulz, T. and Chelaru, A. Model Driven Key Performance Indicators Concepts for Manufacturing Execution Systems. *Applied Mechanics and Materials* 245 (Dec. 2012), pp. 173–178.

[32] JMP. *Correlation*. `https://www.jmp.com/en_us/statistics-knowledge-portal/what-is-correlation.html`. Cited 29/07/2021.

[33] Statistics How To. *Trend Analysis: Simple Definition, Examples*. `https://www.statisticshowto.com/trend-analysis/`. Cited 29/07/2021.

[34] Huspi. *What Is Pattern Recognition in Machine Learning: Complete Guide*. `https://huspi.com/blog-open/pattern-recognition-in-machine-learning/`. Cited 31/07/2021.

[35] Bhattacharya, A. *Effective Approaches for Time Series Anomaly Detection*. `https://towardsdatascience.com/effective-approaches-for-time-series-anomaly-detection-9485b40077f1`. Cited 31/07/2021.

[36] Delling, D., Sanders, P., Schultes, D. and Wagner, D. *Engineering Route Planning Algorithms*. `https://i11www.iti.kit.edu/extra/publications/dssw-erpa-09.pdf`. Cited 23/06/2022.

[37] Sarkar, D. *The Art of Effective Visualization of Multi-dimensional Data*. `https://towardsdatascience.com/the-art-of-effective-visualization-of-multi-dimensional-data-6c7202990c57`. Cited 09/08/2021.

[38] Centers for Disease Control and Prevention. *Using Graphs and Charts to Illustrate Quantitative Data*. `https://www.cdc.gov/healthyyouth/evaluation/pdf/brief12.pdf`. Cited 23/06/2022.

[39] European Commission. *Vehicle categories*. `https://ec.europa.eu/growth/sectors/automotive-industry/vehicle-categories_en`. Cited 13/01/2022.

[40] Britannica. *Railroad history*. `https://www.britannica.com/technology/railroad/Railroad-history`. Cited 13/01/2022.

[41] Huurdeman, A. *The Worldwide History of Telecommunications*. John Wiley & Sons inc., 2003.

[42] VR. *Train Tickets*. `https://www.vr.fi/en/train-tickets`. Cited 30/06/2022.

[43] BusLive. *BusLive*. `https://buslive.eu/`. Cited 30/06/2022.

[44] History. *Automobile history*. `https://www.history.com/topics/inventions/automobiles`. Cited 13/01/2022.

[45] United Nations. *United Nations Conference on Road and Motor transport*. `https://treaties.un.org/doc/Treaties/1952/03/19520326%2003-36%20PM/Ch_XI_B_1_2_3.pdf`. Cited 16/01/2022.

[46] National Museum of American History. *Automobile Safety*. `https://americanhistory.si.edu/america-on-the-move/essays/automobile-safety`. Cited 16/01/2022.

[47] History. *Three-point seatbelt inventor Nils Bohlin born*. `https://www.history.com/this-day-in-history/three-point-seatbelt-inventor-nils-bohlin-born`. Cited 16/01/2022.

[48] Wardell, D. *64 Car Dashboard Warning Lights, Symbols and Meanings*. `https://carinmydna.com/car-dashboard-warning-lights/`. Cited 16/08/2021.

[49] Consumer Reports. *What Does the Check Engine Light Really Mean?* `https://www.consumerreports.org/car-repair-maintenance/what-does-check-engine-light-mean-a2041364753/`. Cited 30/06/2022.

[50] Stafl Systems. *Fully Customizable Touchscreen Displays*. `https://staflsystems.com/touchscreen-displays/`. Cited 30/06/2022.

[51] Moody, L. *Interface concept for Tesla in-car dashboard — a UX case study*. `https://uxdesign.cc/interface-concept-for-tesla-model-3-1b127384b472`. Cited 30/06/2022.

[52] Kusic, L. *How To Make Your Car AUX Volume Louder*. `https://vehiclefreak.com/how-to-make-your-car-aux-volume-louder/`. Cited 30/06/2022.

[53] ACEA. *Vehicles in use Europe*. `https://www.acea.auto/files/report-vehicles-in-use-europe-january-2021-1.pdf`. Cited 29/08/2021.

[54] Cain, T. *Overall UK Auto Industry Sales Figures*. `https://www.goodcarbadcar.net/uk-total-auto-industry-sales-figures/`. Cited 05/09/2021.

[55] SMMT. *Used car sales: Q2 2019*. `https://www.smmt.co.uk/2019/08/used-car-sales-q2-2019/`. Cited 05/09/2021.

[56] Armstrong, M. *Most Important Factors When Buying a Car*. `https://www.statista.com/chart/13075/most-important-factors-when-buying-a-car/`. Cited 30/06/2022.

[57] AutoeBid. *Why mileage matters: the importance of mileage on a used car*. `https://www.autoebid.com/blog/used-cars/importance-of-mileage-on-a-car`. Cited 30/06/2022.

[58] Sutton, M. *How to Test Drive a Car*. `https://www.caranddriver.com/shopping-advice/a15105213/how-to-test-drive-a-car/`. Cited 30/06/2022.

[59] Synopsys. *What is ADAS?* `https://www.synopsys.com/automotive/what-is-adas.html`. Cited 30/06/2022.

[60] Díaz-Lanchas, J., Zofío, J. L. and Llano, C. A trade hierarchy of cities based on transport cost thresholds. *Regional Studies* 0.0 (2021), pp. 1–18. URL: `https://doi.org/10.1080/00343404.2021.1967922`.

[61] Logistiikan Maailma. *Dimensions and Weights in Road Transport*. `https://www.logistiikanmaailma.fi/en/choosing-mode-of-transport/road-transport/dimensions-and-weights/`. Cited 18/01/2022.

[62] International Society for Weigh in Motion. *WIM Systems and Output*. `www.is-wim.org`. Cited 18/01/2022.

[63] Chew, K. F. *Truck Load Capacity Optimisation for Transportation Management System (TMS)*. `https://www.linkedin.com/pulse/truck-load-capacity-optimisation-transportation-system-chew-`. Cited 30/06/2022.

[64] Urban Bus Toolkit. *Factors Influencing Bus System Efficiency*. `https://ppiaf.org/sites/ppiaf.org/files/documents/toolkits/UrbanBusToolkit/assets/1/1d/1d4.html`. Cited 30/06/2022.

[65] Stelzer, A., Englert, F., Horold, S. and Mayas, C. Using customer feedback in public transportation systems. *2014 International Conference on Advanced Logistics and Transport, ICALT 2014* (May 2014), pp. 29–34.

[66] Heide, M. and Gronhaug, K. Respondents' Moods As a Biasing Factor in Surveys: an Experimental Study. *NA - Advances in Consumer Research* 18 (1991), pp. 566–575.

[67] National Strategic Group. *Why Would They Write That?! The Psychology of Customer Reviews*. `https://www.nationalstrategic.com/why-would-they-write-that-the-psychology-of-customer-reviews/`. Cited 02/02/2022.

[68] Morhr, D., Pokotilo, V. and Woetzel, J. *Urban transportation systems of 25 global cities*. `https://www.mckinsey.com/~/media/mckinsey/businessfunctions/operations/ourinsights/buildingatransportsystemthatwork` `elements-of-success-urban-transportation-systems-of-25-global-cities-july-2021.pdf`. Cited 04/02/2022.

[69] Fitzová, H., Matulová, M. and Tomeš, Z. Determinants of urban public transport efficiency: case study of the Czech Republic. *European Transport Research Review* (Sept. 2018).

[70] European Union. *Regulation (EU) 2019/2144 of the European parliament and of the council*. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2144&from=EN`. Cited 17/02/2022.

[71] LeBlanc, D. and Liu, H. *Metrics and Models to Evaluate Driving Safely*. `https://www.nhtsa.gov/sites/nhtsa.gov/files/2021-06/Metrics%20and%20Models%20to%20Evaluate%20Driving%20Safety%20UMTRI%20GI_G104-A.Svenson_0.pdf`. Cited 06/06/2022.

[72] Naderipour, A., Abdul-Malek, Z., Arshad, R. N., Kamyab, H., Chelliapan, S., Ashokkumar, V. and Tavalaei, J. Assessment of carbon footprint from transportation, electricity, water, and waste generation: towards utilisation of renewable energy sources. *Clean Technologies and Environmental Policy* 23 (Jan. 2021), pp. 183–201.

[73] Uber Technologies Inc. *How does Uber work?* `https://www.uber.com/fi/en/about/how-does-uber-work/`. Cited 12/02/2022.

[74] Vervelogic. *The 20 Best Vehicle Tracking Apps for Android Device*. `https://www.vervelogic.com/blog/best-vehicle-tracking-apps-for-android/`. Cited 28/05/2022.

[75] Ludwig, B., Zenker, B. and Schrader, J. Recommendation of Personalized Routes with Public Transport Connections. 53 (Jan. 2009), pp. 97–107.

[76] Helsinki regional transport. *Discounted travel*. `https://www.hsl.fi/en/tickets-and-fares/discounted-travel#lapset`. Cited 20/02/2022.

[77] Helsinki regional transport. *Prams*. `https://www.hsl.fi/en/travelling/accessibility/prams`. Cited 20/02/2022.

[78] Carslaw, D. *Emissions deterioration – the Cinderella of vehicle emissions measurement*. `https://ee.ricardo.com/news/emissions-deterioration-%E2%80%93-the-cinderella-of-vehicl`. Cited 30/06/2022.

[79] Dale, N. *13 Ways to Improve Fuel Economy and Save Money*. `https://www.evanshalshaw.com/blog/13-way-to-improve-fuel-economy-and-save-money/`. Cited 30/06/2022.

[80] Azuga. *What Is Vehicle Fleet Management?* `https://www.azuga.com/fleet-tracking-glossary/vehicle-fleet-management`. Cited 30/05/2022.

[81] Johnson, L. *The complete guide to fleet management*. `https://www.geotab.com/blog/what-is-fleet-management/`. Cited 30/05/2022.

[82] Sun, P. et al. *Scalability in Perception for Autonomous Driving: Waymo Open Dataset*. `https://arxiv.org/pdf/1912.04838v7.pdf`. Cited 28/05/2022.

[83] Kalsbeek, R. *Where to start with the 4 types of analytics*. `https://iterationinsights.com/article/where-to-start-with-the-4-types-of-analytics/`. Cited 26/02/2022.

[84] European Commission. *Financial reporting*. `https://ec.europa.eu/info/business-economy-euro/company-reporting-and-auditing/company-reporting/financial-reporting_en`. Cited 27/02/2022.

[85] Morris, A. *Descriptive Analytics Defined: Benefits & Examples*. `https://www.netsuite.com/portal/resource/articles/erp/descriptive-analytics.shtml`. Cited 28/05/2022.

[86] Holliday, M. *What Is Diagnostic Analytics? How It Works and Examples*. `https://www.netsuite.com/portal/resource/articles/data-warehouse/diagnostic-analytics.shtml`. Cited 28/05/2022.

[87] Maisel, L. and Cokins, G. *Predictive Business Analytics: Forward-Looking Capabilities to Improve Business Performance*. John Wiley & Sons inc., 2014.

[88] Talend. *What Is Prescriptive Analytics?* `https://www.talend.com/resources/what-is-prescriptive-analytics/`. Cited 28/05/2022.

[89] Exeros Technologies. *How Driver Training Will Save Your Business Money*. `https://exeros-technologies.com/how-driver-training-will-save-your-business-money/`. Cited 28/05/2022.

[90] Östlund, J. et al. *Driving performance assessment - methods and metrics*. `http://www.aide-eu.org/pdf/sp2_deliv_new/aide_d2_2_5.pdf`. Cited 28/05/2022.

[91] American Trucking Associations. *The truth about trucking turnover*. `https://www.trucking.org/news-insights/truth-about-trucking-turnover`. Cited 28/05/2022.

[92] NRMA. *Driver training programs for businesses*. `https://www.mynrma.com.au/business/driver-training/programs`. Cited 07/05/2022.

[93] Into training Australia. *Industry Parnerships*. `https://www.intotraining.com.au/services-partnerships/industry-partnerships/`. Cited 07/05/2022.

[94] Decker. *Maximize Your Monthly Bonus Payouts – Monthly Performance Bonus Details 2018*. `https://drivedecker.com/maximize-your-monthly-bonus-payouts-monthly-bonus-details-2018/`. Cited 07/05/2022.

[95] Martin Transportation Systems. *Driver Bonuses*. `https://www.mtstrans.com/driver-bonuses/fuel-economy-bonus-tips/`. Cited 07/05/2022.

[96] Pitcher, M. *Your Driver Performance Plan: 5 things you are (probably) doing wrong*. `https://www.linkedin.com/pulse/your-driver-performance-bonus-5-things-you-probably-doing-pitcher`. Cited 07/05/2022.

[97] Indeed. *Truck Driver salary in United States*. `https://www.indeed.com/career/truck-driver/salaries`. Cited 28/06/2022.

[98] Carter Logistics Incorporated. *Bonus Announcement Letter*. `https://carter-express.com/wp-content/uploads/2019-Bonus-Announcement-Letter.docx.pdf`. Cited 28/06/2022.

[99] Schneider. *Driver pay: Schneider improves performance pay, adds automatic increases*. `https://schneiderjobs.com/blog/improved-truck-driver-performance-pay`. Cited 28/06/2022.

[100] Melton Truck Lines, Inc. *Melton Flatbed Truck Driver PRIDE Program*. `https://drive4melton.com/performance-bonus.php`. Cited 28/06/2022.

[101] Huff, A. *Performance-based pay, Part 1: the science of scoring drivers*. `https://www.ccjdigital.com/business/article/14928600/performance-based-pay-part-1-the-science-of-scoring-drivers`. Cited 28/06/2022.

[102] Lieggi, J. *Improving Pay Packages and Overall Capabilities, H.O.Wolding is Merging with Britton and Bison Transporrt as Bison USA*. `https://www.consumerproductstimes.com/amp/article/561493180-improving-pay-packages-and-overall-capabilities-h-o-wolding-is-merging-with-britton-and-bison-transport-as-bison-usa/`. Cited 28/06/2022.

[103] Prytz, R. *Machine learning methods for vehicle predictive maintenance using off-board and on-board data*. `https://www.diva-portal.org/smash/get/diva2:789498/FULLTEXT01.pdf`. Cited 28/05/2022.

[104] Mind Tools. *Murphy's Law*. `https://www.mindtools.com/pages/article/newSTR_MurphyLaw.htm`. Cited 08/05/2022.

[105] Chaudhuri, A. *Predictive Maintenance for Industrial IoT of Vehicle Fleets using Hierarchical Modified Fuzzy Support Vector Machine*. `https://arxiv.org/ftp/arxiv/papers/1806/1806.09612.pdf`. Cited 08/05/2022.

[106] Lakshmanan, L. *Predictive Maintenance in the Automotive Industry — An Insider's Perspective*. `https://medium.com/embitel-technologies/predictive-maintenance-in-the-automotive-industry-an-insiders-perspective-57388d008c9d`. Cited 08/05/2022.

[107] McFarland, M. *Your smartphone knows if you're a good driver*. `https://money.cnn.com/2016/08/17/technology/smartphone-driver-safety/`. Cited 28/05/2022.

[108] Synopsys. *What is an Autonomous Car?* `https://www.synopsys.com/automotive/what-is-autonomous-car.html`. Cited 13/05/2022.

[109] Allianz Global Corporate & Specialty. *Global Risk Dialogue.* `https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/grd/AGCS-GRD-Fall-Winter-2021.pdf`. Cited 12/05/2022.

[110] McCauley, R. *The 6 Challenges of Autonomous Vehicles and How to Overcome Them.* `https://www.govtech.com/fs/the-6-challenges-of-autonomous-vehicles-and-how-to-overcome-them.html`. Cited 29/06/2022.

[111] Spray, A. *8 Biggest Challenges With Self-Driving Cars.* `https://www.hotcars.com/self-driving-cars-challenges/`. Cited 29/06/2022.

[112] Elliot, L. *Tesla Lawsuit Over Autopilot-Engaged Pedestrian Death Could Disrupt Automated Driving Progress.* `https://www.forbes.com/sites/lanceeliot/2020/05/16/lawsuit-against-tesla-for-autopilot-engaged-pedestrian-death-could-disrupt-full-self-driving-progress/?sh=17fc0d8a71f4`. Cited 13/05/2022.

[113] Ardeen Strategy & Research. *Tech and Legal Challenges the Autonomous Car Industry is Facing.* `https://www.aberdeen.com/blog-posts/tech-legal-challenges-autonomous-car-industry-facing/`. Cited 29/06/2022.

[114] IEEE. *How Will Car Maintenance Evolve In the Autonomous Vehicles Era?* `https://innovationatwork.ieee.org/how-will-car-maintenance-evolve-autonomous-vehicles-era/`. Cited 12/05/2022.

[115] Transport for London. *Accessible Bus Stop Design Guidance.* `https://content.tfl.gov.uk/bus-stop-design-guidance.pdf`. Cited 29/06/2022.

[116] Pinna, I. et al. *Automatic passenger counting systems for public transport.* `https://www.intelligenttransport.com/transport-articles/3116/automatic-passenger-counting-systems-for-public-transport/`. Cited 29/06/2022.

[117] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. *You Only Look Once: Unified, Real-Time Object Detection.* `https://arxiv.org/pdf/1506.02640v5.pdf`. Cited 27/06/2022.

[118] Haider, A., Shaukat, F. and Mir, J. Human detection in aerial thermal imaging using a fully convolutional regression network. *Infrared Physics & Technology* 116 (2021), p. 103796. ISSN: 1350-4495. URL: `https://www.sciencedirect.com/science/article/pii/S1350449521001687`.

[119] Tholen, C. *What Is a Telematics Device?* `https://www.safewise.com/faq/auto-safety/telematics-device/`. Cited 26/06/2022.

[120] IntelliTrac. *Garbage Truck GPS Telematics Solutions.* `https://www.intellitrac.com.au/WasteManagement.html`. Cited 27/06/2022.

[121] European Commission. *What are my rights?* `https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights/what-are-my-rights_en`. Cited 27/06/2022.

[122] Progressive. *Get Snapshot from Progressive*. `https://www.progressive.com/auto/discounts/snapshot/?kbid=62750`. Cited 27/06/2022.

[123] Surya, M. *The Decade of Artificial Intelligence*. `https://towardsdatascience.com/the-decade-of-artificial-intelligence-6fcaf2fae473`. Cited 27/06/2022.

[124] Ford. *FordPass*. `https://www.ford.com/support/category/fordpass/`. Cited 27/06/2022.

[125] Redmon, J. and Farhadi, A. *YOLOv3: An Incremental Improvement*. `https://arxiv.org/abs/1804.02767`. Cited 27/06/2022.

[126] Gupta, A. *Five challenges in designing a fully autonomous system for driverless cars*. `https://www.iiot-world.com/artificial-intelligence-ml/artificial-intelligence/five-challenges-in-designing-a-fully-autonomous-system-for-driverless-cars/`. Cited 29/06/2022.

[127] Barabas, I. and Todurut, A. and Cordos, N. and Molea, A. *Current Challenges in Autonomous Driving*. `https://iopscience.iop.org/article/10.1088/1757-899X/252/1/012096/pdf`. Cited 29/06/2022.